

2005

Application Of Statistical Methods In Risk And Reliability

Astrid Heard
University of Central Florida



Part of the [Mathematics Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Heard, Astrid, "Application Of Statistical Methods In Risk And Reliability" (2005). *Electronic Theses and Dissertations, 2004-2019*. 566.

<https://stars.library.ucf.edu/etd/566>



APPLICATION OF STATISTICAL METHODS IN RISK AND RELIABILITY

by

ASTRID E. HEARD

B.S. University of South Florida, 1972

M.S. Georgia Institute of Technology, 1974

M.S. University of Central Florida, 1992

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
in the College of Arts and Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2005

Major Professor: Dr. Marianna Y. Pensky

© 2005 Astrid E. Heard

ABSTRACT

The dissertation considers construction of confidence intervals for a cumulative distribution function $F(z)$ and its inverse, $F^{-1}(u)$, at some fixed points z and u on the basis of an i.i.d. sample $\underline{X} = \{x_i\}_{i=1}^n$, where the sample size is relatively small. The sample is modeled as having the flexible Generalized Gamma distribution with all three parameters being unknown. This approach can be viewed as an alternative to nonparametric techniques which do not specify distribution of X and lead to less efficient procedures. The confidence intervals are constructed by objective Bayesian methods and use the Jeffreys noninformative prior. Performance of the resulting confidence intervals is studied via Monte Carlo simulations and compared to the performance of nonparametric confidence intervals based on binomial proportion. In addition, techniques for change point detection are analyzed and further evaluated via Monte Carlo simulations. The effect of a change point on the interval estimators is studied both analytically and via Monte Carlo simulations.

This Dissertation is dedicated to my father, Dr. Bernard G. Harvey. His encouragement gave me the confidence to proceed. His knowledge and insight opened my eyes to the possibilities of discovery enabled by mathematics.

ACKNOWLEDGMENTS

In grateful acknowledgement of my family's support and encouragement. I could not have done this without them. Thanks to my mom, Berta Harvey, who supported and encouraged me even though she thought I was crazy to do this. Thanks to Rianna and Michael for believing in their mom. And special thanks to my husband, Sam , who always supports whatever crazy scheme I come up with and kept me from quitting when I had the chance.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	ix
1.0 INTRODUCTION	1
2.0 BACKGROUND AND TECHNICAL APPROACH.....	11
2.1 The Generalized Gamma Distribution.....	12
2.2 Bayesian Analysis.....	18
2.2.1 Bayes Estimation	19
2.2.2 Bayes Hypothesis Testing.....	21
2.3 Bayesian Credible Sets	23
2.4 The Jeffreys Prior.....	25
2.5 Existing Non-Parametric Methods.....	27
2.6 Detection of Time Trend in Data.....	30
2.7 Change Point Analysis.....	32
2.7.1 Formulation of the Problem.....	33
2.7.2 Schwartz Information Criteria (SIC)	34
3.0 TECHNICAL RESULTS.....	37
3.1 Derivation of the Jeffreys Prior.....	39
3.2 The Posterior Distribution of a , b , and λ	42
3.3 The Posterior Distribution of $F(z)$ and $F^{-1}(u)$	44
3.4 Confidence Intervals	47

3.5	Justification of the Approach	49
3.6	Change Point Detection and Location	52
3.6.1	Problem Formulation for Change Point Detection and Location	54
3.6.2	Bayes Factor Change Point Detection and Location	58
3.6.3	Alternative Bayes Technique for Change Point Detection and Location ...	61
3.6.4	Summary	63
3.7	Confidence Intervals In the Presence of a Change Point.....	65
3.7.1	Confidence Intervals for $F(z)$	68
3.7.2	Confidence Intervals for Quantile Z	70
3.8	Error in Confidence Intervals When Change Point is Ignored	71
3.8.1	Confidence Intervals for the CDF.....	72
3.8.2	Confidence Intervals for the Quantile.....	73
4.0	SIMULATIONS	75
4.1	Random Sample Generation	76
4.2	Confidence Intervals Without Change Points.....	77
4.3	Robustness of Technique	82
4.4	Change Point Detection and Location Techniques.....	86
4.5	Confidence Intervals with Change Points.....	93
5.0	CONCLUSIONS.....	99
	LIST OF REFERENCES	102

LIST OF FIGURES

Figure 1: Orbiter Digital Processing System Problem Reports	2
Figure 2: KSC Civil Service Lost Time Injuries	3
Figure 3: Plot of $H(b)$ vs. b	41
Figure 4: Model 1	88
Figure 5: Model 2	88
Figure 6: Model 2	94
Figure 7: Model 2R	94
Figure 8: ε_U vs. k	96
Figure 9: $U, W_k, F_k(z)$ vs. k	97
Figure 10: ε_U vs. k	97
Figure 11: $U, W_k, F_k(z)$ vs. k	98

LIST OF TABLES

Table 1: Change Point Detection/Estimation Techniques	64
Table 2: Lower Confidence Bounds,U, for $u=F(z)$, 1000 Simulations.....	79
Table 3: Upper Confidence Bounds, $\hat{\rho}_u$, for $\rho_u = F^{-1}(u)$, 1000 Simulations.....	80
Table 4: Lower Confidence Bounds,U, for $u=F(z)$, Poisson Data, 1000 Simulations.....	84
Table 5: Upper Confidence Bounds, $\hat{\rho}_u$, for $\rho_u = F^{-1}(u)$, Poisson Data, 1000 Simulations	85
Table 6: Change Point Detection Accuracy (%).....	90
Table 7: Change Point Estimates	92

1.0 INTRODUCTION

At KSC and other NASA Space Flight Operations (SFO) centers, a great deal of effort is expended to collect, analyze and report statistical data on performance of space vehicle systems during tests and operations. In all cases, an effort is made to mitigate the anomaly risk and improve safety and reliability by finding systems that may benefit from some sort of corrective action. Statistical data summarizing performance of space vehicle systems can sometimes enable evaluation of the best possible type of corrective action to use, such as replacement versus redesign. Ultimately, the final decision for vehicle launch is based on the belief that all possible actions have been taken to ensure systems continue to operate safely and reliably.

Figure 1 is a representative plot of problem report counts for the Space Shuttle Orbiter Digital Processing System for a three month time period. Similar plots for all major Space Shuttle subsystems are produced periodically and used to determine the existence of adverse trends requiring additional investigation.

Similarly, safety and health data related to personnel issues is collected in the form of metrics that count occurrences of events of interest. These events are generally relevant to assuring that personnel are operating at optimal health and safety levels necessary for peak performance of their duties and responsibilities, thereby contributing to the overall safety and reliability of a mission. Metric data is frequently evaluated quarterly, based on

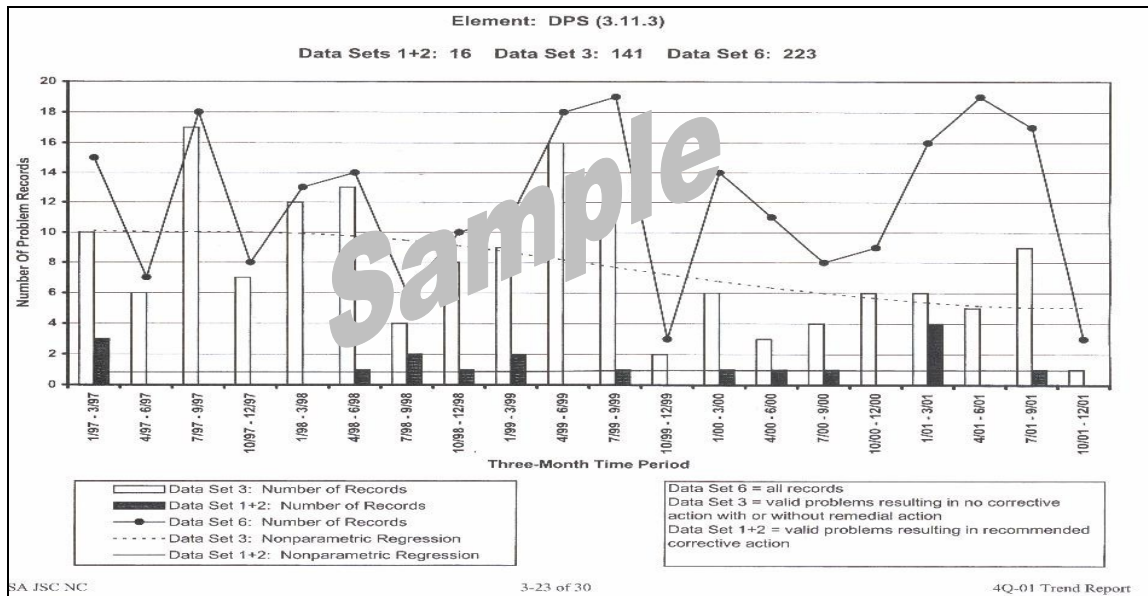


Figure 1: Orbiter Digital Processing System Problem Reports

a limited prior time period, using histograms for visual interpretation of data. Sometimes a “goal” value is displayed, and the data is evaluated based on the relationship of the data to the “goal” value, i.e. whether or not data exceeded (or remained below) this value. Figure 2 is an example of one such set of metric data, the frequency of KSC Civil Service Lost Time Injuries.

While selective engineering analysis does occur, the current practice of initial evaluation of space vehicle systems, as well as assessment of safety and health of the personnel operating them, relies heavily on visual examination of data, usually represented in histogram form, commonly referred to as Pareto charts.

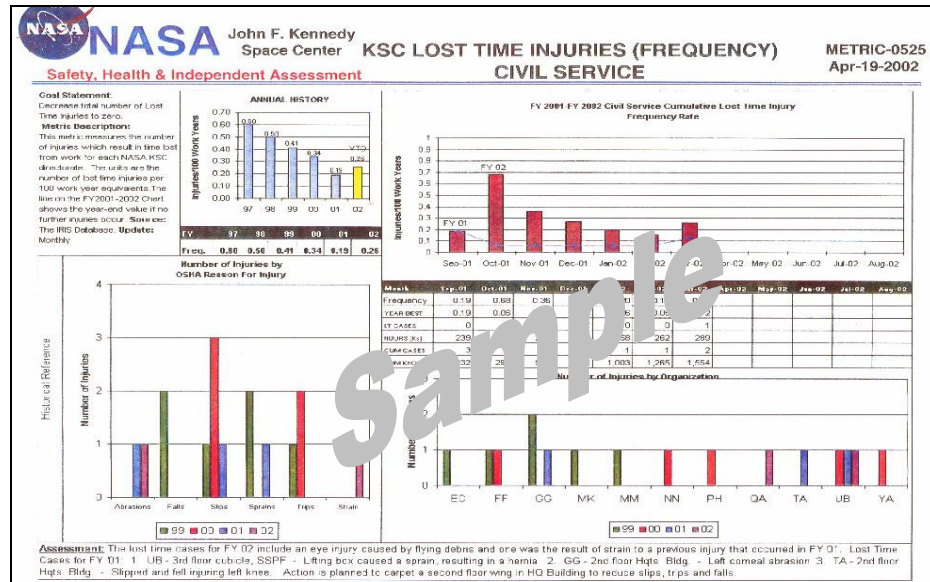


Figure 2: KSC Civil Service Lost Time Injuries

The above examination currently has one objective, namely assessment of the presence of an unfavorable time trend in data for evaluation of whether the system under scrutiny is to be considered acceptable or sufficiently reliable to accomplish a space mission or some other predefined objective. However, it is clear that the absence of trend does not guarantee system safety. It is possible that a system under consideration is unacceptable, even prior to the current period of examination, but did not “fail”, as indicated by data, due to the stochastic nature of the situation.

To make a simple illustration, one may consider someone rolling a couple of fair die and summing the face values. This person may not achieve “twelve” after say, thirty rolls are

performed. But the probability of getting twelve on the next roll remains at $1/36$. There is no guarantee that twelve will not be achieved on the thirty-first roll. A similar situation occurs in reliability. If there has not been a significant anomaly, there is no guarantee an anomaly will not occur in the future just because it has not happened before.

This demonstrates that it is vital to assess reliability of parts and systems of space vehicles even if time trend is absent. Initially, data can be examined to establish the presence of trend. However in many situations, as in the example we considered above, time trend is not present due to the identical distribution of the outcomes of the rolls. In much the same way, the problem counts derived from the NASA/KSC Space Shuttle Problem Reporting and Corrective Action (PRACA) database are very often trend free, and for this reason are not examined from a reliability point of view under the present system.

The goal of this dissertation is to study trend-free data from the point of view of reliability. In order to draw a data-driven decision, program managers usually want information on what is the probability that the number of problem counts X of a certain system does not exceed a given threshold (say, z), and the program manager wants to be, say, $\beta * 100\%$ confident in the results. This, in fact, is a very common problem since the probability that a random variable of interest, X , does not exceed a certain pre-specified value $z : P(X \leq z) = F(z)$, the distribution function of variable X at the point z , is the main object of inference in many practical applications.

For example, if X is a lifetime of a device under scrutiny, $1-F(z)$ is just the probability that this device will function for a longer time than z . If X is the number of failures of a certain equipment within a fixed time period, then $F(z)$ is the probability that the number of failures within a given time period is not more than z , and so on.

In practice, however, a point estimator of $F(z)$ makes little sense since the actual value of $F(z)$ may vary from its estimator quite significantly if the variance of the estimator is high. As a consequence, the information which decision making personnel really want is a lower bound on $F(z)$, i.e. the value U such that, say, $P(F(z) > U) > 99\%$. Sometimes, there is a need in solution of the inverse problem, namely, for a given probability u find a number Z such that $P(F(Z) > u) > 99\%$. The former can be re-written as $P(Q(u) < Z) > 99\%$ where $Q(u) = F^{-1}(u)$ is the quantile function of X at the point u .

In more precise terms, these problems can be formulated as follows. Let $\underline{X} = \{x_i\}_{i=1}^n$ be an iid sample of a random variable X where the values x_i represent the number or frequencies of occurrences of some events (e.g. system anomalies) within given increments of time. Let this sample have an unknown density function f and cumulative distribution function F . Let $Q(u) = F^{-1}(u)$ be the quantile function of X . The questions above can be posed as:

(Q1) Given iid sample \underline{X} of a random variable X and the values $z \in (0, \infty)$ and $\beta \in (0, 1)$, find $U = U(\underline{X}) \in (0, 1)$ such that

$$P(P(X \leq z) \geq U) = P(F(z) \geq U) = \beta, \quad (1.1)$$

(Q2) Given iid sample \underline{X} of a random variable X and the values $u, \beta \in (0,1)$, find $Z = Z(\underline{X}) \in (0, \infty)$ such that

$$P(F(z) \geq u) = P(Q(u) < Z) = \beta. \quad (1.2)$$

Observe that (Q1) is the problem of construction of the $\beta*100\%$ lower confidence bound for the distribution function F at the known point z while (Q2) is equivalent to construction of the upper confidence bound for its inverse Q at some point u . The first problem is common in guaranteed coverage tolerance prediction described in, for example, Aitchison and Dunsmore [42]. The main challenge here is development of fully data-based techniques of construction of the confidence bounds in (1.1) and (1.2) which are suitable for small sample sizes ($n = 20-50$). In addition, in reliability one is interested in "extreme" values of z and u , namely the values of z such that $F(z) = 1 - \varepsilon$ or $u = 1 - \varepsilon$ where ε is small. Hence, the inference deals with the tails of the distribution F and is based on a very small number of observations. This circumstance rules out nonparametric approaches that cannot successfully deal with the tails of an unknown distribution for small sample sizes. Additionally, nonparametric approaches usually result in relatively long confidence intervals since they are always designed for the "worst case scenario". Hence, we need to choose a class of distributions F that is flexible enough, but can still be treated parametrically.

The distribution families most commonly used to model the lifetimes are the Weibull or gamma distributions. Since they have different functional forms, the practitioners either choose one of them arbitrarily or test the hypothesis which of the two agrees better with the observations (see e.g. Volodin [43]). After distinguishing between the two, and matching the shape parameter to data, the confidence intervals are constructed for the values of $F(z)$ and $Q(u)$. In order to avoid making this choice and design completely data-driven confidence intervals, we assume that the sample \underline{X} is drawn from the generalized gamma distribution introduced by Stacy [4] with the pdf

$$p(x | a, b, \lambda) = \lambda [\Gamma(b)]^{-1} a^{-b} x^{\lambda b - 1} \exp(-x^\lambda / a) \quad (1.3)$$

The advantage of using (1.3) is that it can emulate a wide variety of curves, so that the majority of distributions used in reliability are particular cases of (1.3). Parameter λ defines how thin the tails of the distribution (1.3) are, hence, by accommodating λ close to zero, we include distributions with relatively heavy tails while large values of λ lead to distributions with very thin tails.

It should be noted that due to its flexibility, statistical inference for generalized gamma distribution is rather tricky. For example, it is not uncommon even with the sample size of two to three hundred for algorithms for determining MLE's of a , b and λ to fail to converge (see e.g. Hager and Bain [44]). The way out of this difficulty is either using the rather unreliable method of moments estimators (see e.g. Stacy and Mihram [37]) or assuming one or two parameters of generalized gamma distribution to be known (see e.g. Bain and Weeks [45], Lawless [46] or Pham and Almhana [47]). In practice, however,

parameters of distributions are never known, so they are estimated from the data and then are treated as known. If we were to adopt this approach in the present paper, this would reduce complexity of the problem, however, it would make constructed confidence intervals unreliable, especially since the amount of data available is small.

In order to construct data-driven solution to the questions **(Q1)** and **(Q2)** based on small samples, we shall design Bayesian confidence intervals based on the non-informative Jeffreys prior. In our approach, we shall treat all three parameters of the generalized gamma distribution as completely unknown just making a (quite insignificant) assumption that the shape parameter λ is bounded from above.

After elaborating on the theoretical aspects of construction of the interval estimators, we shall study them thoroughly via Monte Carlo simulations. We shall compare the estimators designed on the basis of parametric assumptions with nonparametric interval estimators as discussed by Brown, Cai, and DasGupta[7] and Cai[6].

The concept of testing for identically distributed sample values will be explored by assessing if there is a change in the distribution of the data. This is called the change point analysis of the sample. In order to accomplish this goal, we shall analyze whether the sample, $\underline{X} = \{x_i\}_{i=1}^n$, is indeed i.i.d. or if for some $1 < k < n$ $\underline{X}_1 = \{x_i\}_{i=1}^k$ has a pdf $f_1(x)$ while $\underline{X}_2 = \{x_i\}_{i=k+1}^n$ has a pdf $f_2(x)$ where $f_1(x) \neq f_2(x)$. This is change point detection and location. Taking the answer to this question into account, we shall then construct

confidence bounds (1.1) and (1.2), and determine the impact of using confidence intervals that ignore a change point versus confidence intervals based on a detected change point. Monte Carlo simulations will explore some sample model outcomes that represent the theoretical results of this change point analysis.

The dissertation is organized as follows. In Chapter 2 we provide background information on the concepts and techniques employed in this proposal. We start with properties of the Generalized Gamma distribution (Section 2.1), then discuss the main ideas of Bayesian analysis (Section 2.2) and construction of Bayesian credible sets (Section 2.3). Section 2.4 explains the Jeffreys non-informative prior and Section 2.5 describes existing techniques for construction of nonparametric confidence intervals for the cdf $F(z)$. Section 2.6 discusses the χ^2 goodness-of-fit test for detection of trend in data. Finally, Section 2.7 provides the framework for the problem of change points in distributions.

Chapter 3 develops the theoretical approach to construction of Bayesian confidence bounds for the cdf $F(z)$ and its inverse, $F^{-1}(u)$, based on Jeffreys non-informative prior. We start with derivation of the Jeffreys prior for the Generalize Gamma distribution (Section 3.1), then discuss the use of the Jeffreys prior in the derivation of the posterior distribution for the three parameters of the Generalized Gamma distribution (Section 3.2). Derivation of the posterior distribution for $F(z)$ and $F^{-1}(u)$ is discussed in Section 3.3 and the computational technique for deriving confidence bounds from these posterior

distributions is discussed in Section 3.4. Section 3.5 justifies the proposed approach by showing that the Jeffreys prior leads to a proper posterior distribution. Finally, Sections 3.6, 3.7, and 3.8 are devoted to developing techniques for change point detection and location, construction of confidence intervals in a view of an existing change point , and evaluating the error which occurs when the change point is ignored.

Chapter 4 considers computer simulations designed to assess the properties of the interval estimators constructed in Chapter 3 and to compare them with the nonparametric estimators described in Section 2.5. This chapter presents simulation results and discusses them in detail. In addition, numerical simulations are used to assess precision of techniques for detection and location of the change point as well as interval estimation with and without taking change point into account.

Chapter 5 provides some general conclusions and discussion of simulation results.

2.0 BACKGROUND AND TECHNICAL APPROACH

This chapter provides background information and explains the technical approach to resolve the problems described in Chapter 1, Introduction. In order to construct β -level one-sided confidence intervals for an unknown distribution function $F(z)$ and its inverse $F^{-1}(u)$, we assume that the sample $\underline{X} = \{x_i\}_{i=1}^n$ has the Generalized Gamma distribution. We discuss properties of the Generalized Gamma distribution in Section 2.1. The confidence intervals mentioned above will be constructed on the basis of a Bayesian approach, hence we briefly consider ideas of Bayesian inference and construction of Bayesian credible sets in Sections 2.2 and 2.3. Since we have no prior knowledge about possible values of unknown parameters of the Generalized Gamma distribution, we shall employ the noninformative Jeffreys prior for those parameters. Section 2.4 contains background information on the Jeffreys noninformative prior.

In order to evaluate Bayesian confidence intervals constructed in this paper, we shall compare them to nonparametric confidence intervals. These intervals are based on the fact that estimation of $F(z)$ at a specific point z is equivalent to estimation of an unknown probability from binomial data. Construction of confidence intervals for binomial probability has been a topic of investigation carried out recently by Cai [6] and Brown et al. [7]. We consider those methods in Section 2.5.

Finally, as discussed briefly in Chapter 1, there is an assumption that the observations x_i are identically distributed with no existing time trend. These assumptions must be justified prior to using the proposed method. Detection of trend in data is discussed in Section 2.6. The issue of distribution being identically distributed versus the existence of a change point is discussed in Section 2.7.

Through simulation, confidence intervals derived by proposed new methods will be compared to nonparametric confidence intervals in terms of coverage probability and average interval length. In addition, the impact of the change point in the sample on confidence intervals is also explored through simulation.

2.1 The Generalized Gamma Distribution

For the parametric problem, the first decision is to select a density function for the assumed distribution of the i.i.d. observations $\underline{X} = \{x_i\}_{i=1}^n$. The Generalized Gamma distribution was selected because, as will be shown, the distribution provides flexibility to include virtually every distribution used in reliability and survival analysis. The only exception may be discrete data that is often represented by a Poisson process. This issue is addressed in [36] to show that in fact the Generalized Gamma distribution can effectively represent a Poisson variable for purposes of confidence interval generation. This provides freedom from the “curse of discreteness” which is reported in relation to binomial data (see Agresti and Coull [27] and Brown, et. al. [7]).

The Generalized Gamma distribution is derived from the well-known gamma distribution for a non-negative random variable X with positive parameters. The probability density function of the Gamma distribution is of the form

$$f(x | \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad (2.1.1)$$

as defined in most textbooks [1].

The specific form of a generalization of the gamma distribution was suggested by Liouville's extension to Dirichlet's integral formula [11]. In this form it also may be regarded as a special case of a function introduced by L. Amoroso [10] and R. d'Addario [11] in analyzing the distribution of economic income. In essence, the generalization (2.1.4) herein is accomplished by supplying a positive parameter, p , as an exponent in the exponential factor of the gamma distribution. It is shown that cumulative probabilities are related directly to the incomplete gamma function (tabulated in [14]).

In his paper, Stacy [4] defines a random variable with the probability density function

$$f(x | a, d, p) = \frac{\left(\frac{p}{a^d}\right) x^{d-1} e^{-(x/a)^p}}{\Gamma\left(\frac{d}{p}\right)} \quad (2.1.2)$$

for non-negative values of x and positive values of the parameters a , d , and p . Stacy [4] refers to this function as the generalization of the gamma distribution. The derivation of this generalized form is obtained using the gamma distribution as defined in (2.1.1) and

applying the transformation of X to a new random variable z^p . Then a substitution for the parameters α and β is performed. This process is described below.

If random variables X and z are related as $X = h(z)$, then the pdf of z can be obtained from the pdf of X using the common transformation formula

$$f(z | \alpha, \beta) = f(h(z) | \alpha, \beta) \left| \frac{dx}{dz} \right|. \quad (2.1.3)$$

If $h(z) = z^p$, then substituting (2.1.1) into (2.1.3) we derive

$$f(z | \alpha, \beta) = \frac{p}{\Gamma(\alpha)\beta^\alpha} z^{\alpha p - 1} e^{-z^p/\beta}.$$

Setting $\alpha = \frac{d}{p}$ and $\beta = a^p$, we obtain equation (2.1.2).

If, without loss of generality, we define new parameters as $a' = a^p$, $b' = \frac{d}{p}$, $\lambda = p$,

then (2.1.2) reduces to

$$f(x | a', b', \lambda) = \frac{\left(\lambda/a'^{b'}\right) x^{\lambda b' - 1} e^{-\left(x^{\lambda}/a'\right)}}{\Gamma(b')} = \frac{\lambda}{\Gamma(b') a'^{b'}} x^{\lambda b' - 1} e^{-\left(x^{\lambda}/a'\right)}.$$

For the remainder of this dissertation, we drop the “prime” in the notation of the parameters to yield the form used hereafter:

$$f_g(x | a, b, \lambda) = \frac{\lambda}{\Gamma(b) a^b} x^{\lambda b - 1} e^{-\left(x^{\lambda}/a\right)}. \quad (2.1.4)$$

Now, let us examine the properties of the Generalized Gamma distribution (2.1.4). The main advantage of the application of this distribution family is that it includes the majority of distribution families used in reliability and survival analysis. This flexibility is due to an extra parameter, λ . In what follows, we explain how the exponential, Weibull, half-normal(without a location parameter), Gamma, Chi-Squared and Rayleigh pdf's can be obtained from representation (2.1.4).

The exponential distribution has the pdf given by

$$f(x | \beta) = \beta e^{-\beta x}. \quad (2.1.5)$$

Setting $\lambda=1$, $b=1$, and $a=\frac{1}{\beta}$, we reduce (2.1.4) to (2.1.5).

The pdf of the Weibull distribution is given by

$$f(x | \lambda, a) = \frac{\lambda}{a} x^{\lambda-1} e^{-\frac{x^\lambda}{a}}. \quad (2.1.6)$$

Setting $b = 1$ in formula (2.1.4) we reduce it to (2.1.6).

The Rayleigh distribution is the Weibull distribution with the shape parameter, $a=2$. So clearly, (2.1.4) can also be reduced to a Rayleigh distribution.

The pdf of the half-normal distribution, centered at zero is given by

$$f(x | 0, \sigma^2) = \frac{2}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad x > 0 \quad (2.1.7)$$

Setting $\lambda=2$, $b=\frac{1}{2}$, and $a=2\sigma^2$ in formula (2.1.4), we reduce it to (2.1.7).

The pdf of the Gamma distribution is given by (2.1.1). Setting $a=\beta$, $b=\alpha$ and $\lambda=1$ in formula (2.1.4) we reduce it to (2.1.1)

The pdf of the Chi-Square distribution is given by

$$f(x|p) = \frac{1}{\Gamma\left(\frac{p}{2}\right)2^{p/2}} x^{(p/2)-1} e^{-\frac{x}{2}}. \quad (2.1.8)$$

Setting $b = \frac{p}{2}$, $a=2$, and $\lambda=1$ in formula (2.1.4) we reduce it to (2.1.8).

The above shows the flexibility of the Generalized Gamma distribution. This leads us to the conclusion that the Generalized Gamma distribution provides a model flexible enough for the distribution of the random sample $\underline{X} = \{x_i\}_{i=1}^n$. Thus, we use this distribution for developing a parametric approach to construction of confidence intervals.

Denote the cumulative distribution function of the Generalized Gamma distribution by

$F_z(x|a,b,\lambda)$ and the Incomplete Gamma function by $\Gamma(t,z) = \int_z^\infty s^{t-1} e^{-s} ds$. Integrating

(2.1.4) with respect to x yields

$$F_z(x|a, b, \lambda) = F(z) = \int_0^z f(x|a, b, \lambda) dx = 1 - \frac{\Gamma\left(b, \frac{z^\lambda}{a}\right)}{\Gamma(b)}. \quad (2.1.9)$$

Define the Regularized Incomplete Gamma Function

$$Q[a, t] = \frac{\Gamma[a, t]}{\Gamma[a]} = \frac{1}{\Gamma[a]} \int_t^\infty s^{a-1} e^{-s} ds = \xi \quad (2.1.10)$$

and the Regularized Inverse Gamma Function, $Q^{-1}[a, \xi]$, by the equation $Q^{-1}[a, \xi] = t$ such that

$$Q^{-1}(a, Q(a, t)) = t.$$

Using (2.1.10), formula (2.1.) can be rewritten as

$$F(z) = 1 - Q\left[b, \frac{z^\lambda}{a}\right] = u \quad (2.1.11)$$

for some $u \in [0, 1]$. Mathematica version 5.1 software contains the Regularized Inverse Gamma Function, enabling us to calculate $z = F^{-1}(u)$. Note that (2.1.11) implies

$$Q\left[b, \frac{z^\lambda}{a}\right] = 1 - u \text{ yielding } Q^{-1}[b, 1 - u] = \frac{z^\lambda}{a} = \frac{(F^{-1}(u))^\lambda}{a}. \text{ Hence } z = F^{-1}(u) \text{ can be}$$

written as

$$F^{-1}(u) = \left(a Q^{-1}[b, 1 - u]\right)^{1/\lambda} = z. \quad (2.1.12)$$

Equations (2.1.11) and (2.1.12) will be used later for calculation of the cumulative distribution function of the Generalized Gamma distribution and its inverse. For a fixed value of $z \in \Re$ and $u \in [0,1]$, confidence intervals for both $F(z)$ and $F^{-1}(u)$ will be computed using Bayesian techniques as described in the next section.

2.2 Bayesian Analysis

Bayesian analysis differs from other statistical approaches in its use of conditional probability as the main tool to update information about the model. In order to understand the Bayesian approach, we begin with a basic definition of conditional probability [1].

Definition: If A and B are events in S, and $P(B) > 0$, then the *conditional probability* of A given that B has occurred, written $P(A|B)$, is:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

which can be rewritten as $P(A \cap B) = P(A | B)P(B)$.

Using this definition of conditional probability, we see that $P(B | A) = \frac{P(A \cap B)}{P(A)}$ which

can be written as $P(A \cap B) = P(B | A)P(A)$. Combining these results for $P(A \cap B)$

yields $P(A|B) = P(B|A) \frac{P(A)}{P(B)}$ that is often called the Bayes rule after the discoverer,

Sir Thomas Bayes.

The Bayes rule has a more general form that applies to partitions of a discrete sample space [1]. For the discrete case, let $\{A_i\}_{i=1}^n$ be the sample space partition, where n can be infinite, and let B be any set. Then for each $i=1,2,\dots,n$ we have

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}. \quad (2.2.1)$$

For a continuous sample space, we consider a bivariate random vector (X, θ) where X represents the sample vector and θ represents the vector of parameters with the joint density function $\pi(X, \theta)$. Then the marginal distribution of X is $f_X(x) = \int \pi(x, \theta) d\theta$ and the marginal distribution of θ is $\pi_\theta(\theta) = \int \pi(x, \theta) dx$, also called the *prior* density function of θ .

2.2.1 Bayes Estimation

Usually, it is assumed that a random variable (or sample) X has a known density function $f_X(x|\theta)$, containing an unknown parameter θ . In Bayesian analysis, it is assumed that θ is a random variable with the $\pi_\theta(\theta) > 0$, hence the probability density $f_X(x|\theta)$ is

viewed as a conditional density of X given θ . After the actual value of X is observed, the information about the density of θ is updated using an analog of formula (2.2.1)

$$f_X(x|\theta) = P(X|\theta) = \frac{\pi(X, \theta)}{\pi_\Theta(\theta)} \quad \text{and} \quad P(\theta|X) = \frac{\pi(X, \theta)}{f_X(x)}.$$

Combination of the above equations yields the formula for computation of the posterior density function of parameter vector, θ , given a random sample, \underline{X}

$$\pi_p(\theta) = P(\theta|X) = \frac{P(X|\theta)\pi_\Theta(\theta)}{f_X(x)}. \quad (2.2.2)$$

Here and in what follows we use the subscript p to label (2.2.2) as the posterior distribution dropping explicit indication of its dependence on the sample \underline{X} . Thus,

$$\pi_p(\theta) = \frac{P(X|\theta)\pi_\Theta(\theta)}{f_X(x)} = \frac{f_X(X|\theta)\pi_\Theta(\theta)}{\int \pi(x, \theta) d\theta} = \frac{f_X(X|\theta)\pi_\Theta(\theta)}{\int f_X(X|\theta)\pi_\Theta(\theta) d\theta}. \quad (2.2.3)$$

This posterior density function for the three parameters of the Generalized Gamma distribution will be derived in Chapter 3.

Our objective is to apply Bayesian techniques for derivation of the posterior density of $F(z)$ for a fixed value of z and the density of its inverse, $F^{-1}(u)$, at a fixed points z and u , where F is the cumulative distribution function of the Generalized Gamma distribution. These posterior density functions will then be used for construction of confidence intervals. Since no specific prior information on the parameters of the Generalized Gamma distribution is available, the noninformative Jeffreys' prior, will be used. This

selection of a prior, and its justification is discussed in Section 2.4. The application of the Bayes technique to derivation of the posterior density of $F(z)$ and its inverse, $F^{-1}(u)$, is discussed in Section 3.

2.2.2 Bayes Hypothesis Testing

Our second objective is to apply Bayesian techniques for testing a hypothesis about the existence of a change point in the random sample, as discussed in Section 2.7. The general decision on whether to accept or reject a null hypothesis is based on the Bayes factor (BF), which is the ratio between the posterior probability of the null hypothesis and the alternative hypothesis, respectively given the observations, where the posterior probabilities are calculated on the basis of Bayes rule:

$$BF = \frac{P(H_0 | \underline{X})}{P(H_1 | \underline{X})} = \frac{P(H_0, \underline{X})}{P(H_1, \underline{X})} . \quad (2.2.4)$$

The null hypothesis, H_0 , is accepted if $BF \geq L$ and rejected if $BF < L$ where $L > 0$ is a threshold which is specified in advance. As a general rule, $L=1$. We set $L < 1$ only if stronger evidence of the alternative hypothesis is required for acceptance, and $L > 1$ if stronger evidence of the null hypothesis is required for acceptance.

If A and B are disjoint sets, the null hypothesis tests $\theta \in A$ versus $\theta \in A$. For parameter θ prior distributions are assigned corresponding to the null and alternative hypothesis, say $\pi(\theta | \underline{\alpha})$ and $\pi(\theta | \underline{\beta})$ where $\underline{\alpha}$ and $\underline{\beta}$ are scalar or vector parameters. Parameters

$\underline{\alpha}$ and $\underline{\beta}$ may be equal or one may be a subset of the other. Then the Bayes Factor in (2.2.4) becomes:

$$BF = \frac{P(H_0, \underline{X})}{P(H_1, \underline{X})} = \frac{\int_{H_0: \theta \in A} P(\underline{X} | \theta) \pi(\theta | \alpha) d\theta}{\int_{H_1: \theta \in B} P(\underline{X} | \theta) \pi(\theta | \beta) d\theta} \quad (2.2.5)$$

Therefore, in order to determine the value of the Bayes Factor, the values of $\underline{\alpha}$ and $\underline{\beta}$ must be estimated. If $\underline{\alpha} = \underline{\beta} = p$ is a common parameter, the value of p can be estimated by maximizing the likelihood:

$$P(p | \underline{X}) \sim P(\underline{X} | p) = \int_{\theta \in A \cup B} P(\underline{X} | \theta) \pi(\theta | p) d\theta$$

where $A \cup B$ are all possible values of θ for both the null and alternative hypothesis.

Hence, the estimator of p is of the form $\hat{p} = \arg \max P(\underline{X} | p)$ where maximum is evaluated over all possible values of p .

If a sensible maximum cannot be found, then a hierarchical Bayes approach can be used.

Hierarchical Bayes approach assumes that p itself is a random variable with pdf $h(p | \underline{\phi})$

where $\underline{\phi}$ is a vector of unknown parameters.

Then,

$$BF = \frac{P(H_0, \underline{X})}{P(H_1, \underline{X})} = \frac{\iint_{H_0: \theta \in A, p \in C} P(\underline{X} | \theta) \pi(\theta | p) h(p | \phi) d\theta dp}{\iint_{H_1: \theta \in B, p \in C} P(\underline{X} | \theta) \pi(\theta | p) h(p | \phi) d\theta dp} \quad (2.2.6)$$

and ϕ can be estimated by maximizing:

$$P(\underline{X} | p) = \iint_{\theta \in A \cup B, p \in C} P(\underline{X} | \theta) \pi(\theta | p) h(p | \phi) d\theta dp \quad (2.2.7)$$

The application of these Bayes hypothesis testing techniques to change point detection and estimation are developed in Section 3.

2.3 Bayesian Credible Sets

An interval estimate of a real-valued parameter θ based on a random sample $\underline{X} = \{x_i\}_{i=1}^n$ is the random interval $[L(\underline{X}), U(\underline{X})]$ where $L(\underline{X}) \leq U(\underline{X})$. The parameter θ is covered by the interval $[L(\underline{X}), U(\underline{X})]$ with some probability β that is called the confidence level. The interval estimator $[L(\underline{X}), U(\underline{X})]$, together with a measure of confidence, is known as a *confidence interval or a set*.

Within classical statistics, parameters of distributions are assumed to be fixed unknown constants. A $(1 - \alpha)$ confidence interval $[L(\underline{X}), U(\underline{X})]$ represents one of the possible realized values of a random interval whose boundaries are defined by the random sample.

Since the parameters are fixed, a parameter is in the given confidence set with probability either 0 or 1. Thus the realized interval has a $100(1-\alpha)\%$ chance of coverage of the true parameter value[1].

Under a Bayesian model, parameters are treated as random variables which have a probability distribution, called a *prior distribution*. To derive a posterior distribution given the prior, one has to use formula (2.2.3), which in the case of the Generalized Gamma function distribution with the pdf (2.1.4), becomes

$$\pi_p(a,b,\lambda) = \frac{f_g(\underline{X}|a,b,\lambda) \pi_x(a,b,\lambda)}{\iiint_{a,b,\lambda} f_g(\underline{X}|a,b,\lambda) \pi_x(a,b,\lambda) da d\lambda db} . \quad (2.3.1)$$

Here $\pi_x(a,b,\lambda)$ is the prior distribution of the parameters based on a random sample, \underline{X} .

All Bayesian claims of coverage by a confidence set are made with respect to the posterior distribution of the parameters [1]. Thus the Bayesian setup allows the statement that a parameter is inside the confidence set with some probability $(1-\alpha)$, *not* 0 or 1. To keep the distinction between classical confidence sets and those defined using Bayesian techniques, the term *Bayesian credible sets* is used. It is an important distinction because the confidence and the credible sets allow quite different probability assessments about a distribution parameter. By utilizing Bayesian techniques for generating confidence intervals (sets), this proposal will rely on the inferences based on Bayesian credible sets.

2.4 The Jeffreys Prior

In order to use these Bayesian techniques, one must decide on a reasonable prior distribution for the parameters. If no knowledge about these parameters is available, it is sensible to choose a prior that requires minimal assumptions about the parameter values. Thus, a *noninformative* prior will be used.

Because of the compelling reasons to perform a conditional analysis and the attractiveness of using Bayesian techniques to do so, there have been attempts to use the Bayesian approach even when no (or minimal) information is available [3]. A *noninformative prior* is a prior distribution for θ which does not favor any possible value of θ over other values. For example, in testing between two simple hypotheses, the prior which gives probability 1/2 to each of the hypotheses is clearly noninformative.

The uniform density given by $\pi(\theta) = 1$ was introduced and used by Laplace in 1812 [15]. Since the probability of any parameter value is 1, this is a noninformative prior and it has infinite mass. Distributions like this are called *improper* and the Laplace prior shares this characteristic with many other noninformative priors.

The simplest situation to consider is when θ is a finite set consisting of n elements. The obvious noninformative prior is $\pi(\theta) = \frac{1}{n}$. This was routinely done by Laplace and was

criticized because of a lack of invariance under transformation [16]. Efforts to derive noninformative priors which are invariant under transformations of the parameters of the model were started by Jeffreys[8]. The Jeffreys prior has been extensively used ever since. Liseo[5] claims that this is now the most widely used method for determining a noninformative prior.

Jeffreys[8] proposed the use of

$$\pi_J(\omega) = |\det H(\omega)|^{\frac{1}{2}} \quad (2.3.2)$$

where ω is the vector of parameters and $H(\omega)$ is the expected value of the Fisher information matrix. Under commonly satisfied assumptions [26], the expectation of the Fisher information matrix for the p parameters of the distribution yielding the vector $\omega = \{\omega_k\}_{k=1}^p$, is defined by $H(\omega) = \{H_{i,j}\}$ where

$$H_{i,j} = -E \left(\frac{\partial^2}{\partial \omega_i \partial \omega_j} \log \prod_{k=1}^n f(x_k | \omega) \right) .$$

The merit of the *Jeffreys prior* is that it remains invariant under any one to one reparametrization and does not depend on a declared parameter of interest. If the additional assumption is made that the parameters are independent, the Jeffreys prior for the combination of parameters is the product of the noninformative priors obtained separately for each parameter. However, this assumption of parameter independence was *not* made for theory developed in this proposal.

Note that the Jeffreys priors are often improper, i.e. integration with respect to all of the parameters yields infinite value. This is not an impediment to Bayesian analysis as long as the resulting posterior distribution is proper. In the case of the Generalized Gamma distribution, we shall show that the Jeffreys prior leads to a proper posterior distribution.

Armed with these Bayesian techniques, we shall derive a posterior distribution for the parameters of the Generalized Gamma Distribution (2.2.1). After that we shall introduce new random variables $F(z)$ and $F^{-1}(u)$ where z and u are fixed constants. We shall derive their posterior distributions, using transformations of random variables.

Once joint distributions for the newly introduced random variables and the other two parameters are found, the posterior densities of variables $F(z)$ and $F^{-1}(u)$ are derived by integrating out the two nuisance parameters. These posterior distributions, together with the confidence level $\beta=.95$, will be used to construct confidence intervals for $F(z)$ and $F^{-1}(u)$.

2.5 Existing Non-Parametric Methods

Construction of a nonparametric lower confidence bound for $F(z)$ at a known point z is intimately related to interval estimation of a binomial probability p on the basis of n i.i.d. observations. To see this, form a new i.i.d. sample $\underline{Y} = \{Y_i\}_{i=1}^n$ with $Y_i = I(X_i \leq z)$ where $I(A)$ is an indicator function of the set A . Then Y_i are Bernoulli variables with

$p = P(Y_i = 1) = F(z)$, hence $Y = \sum_{i=1}^n Y_i$ has binomial distribution with parameters n and p , and problem **(Q1)** reduces to construction of the $100 * \beta\%$ lower confidence bound for p on the basis of Y . The point estimator of p is $\hat{p} = \frac{Y}{n}$ and the interval estimator is based on \hat{p} .

This problem has a long history and extensive literature coverage (see e.g. Agresti and Coull [27], Blyth and Still [28], Brown[33], Cai[6] and Cressie[29]). The most common 2-sided confidence interval for p is known as the standard or Wald interval as it comes from the Wald large sample test for the binomial parameter p [7]. Since $E(\hat{p}) = p$ and $Var(\hat{p}) = n^{-1}p(1-p)$, this textbook $100 * \beta\%$ lower bound based on asymptotic normality of statistic \hat{p} . and has the form $p_\beta = \hat{p} - \kappa n^{-1/2} (\hat{p}(1-\hat{p}))^{1/2}$ where $\kappa = z_\beta$ is the $100 * (1-\beta)th$ percentile of the standard normal distribution. A modification of this bound (leading to the Wilson, or score interval) is the positive solution of the simple quadratic equation $\hat{p} - p_\beta = \kappa n^{-1/2} (\hat{p}_\beta(1-\hat{p}_\beta))^{1/2}$.

The intervals $[p_\beta, 1]$ listed above, however do not have adequate coverage, especially when p near 0 is close to zero or one (see for example, Cressie[29], Blyth and Still [28], Vollset [32], Santner [31], Agresti and Coull [27] and Newcombe [30]), which are the only cases of interest in reliability. The difficulty is usually attributed to the use of a

large sample approximation when n is small. In their article, Brown, Cai and DasGupta [7] give a comprehensive treatment of two-sided confidence intervals for a binomial proportion. They show that deficiencies of the Wald interval are not confined to just p near zero or one and to small n only. Cai [6] provides examples showing that the one-sided Wald interval suffers a pronounced systematic bias in coverage probability and expected length.

In his paper, Cai resolutely recommends the one-sided Jeffreys and second-order corrected intervals as alternative Wald and score intervals. The Jeffreys interval is the highest posterior density (HPD) interval constructed by Bayesian inference with Jeffreys prior which in this case is $\text{Beta}(0.5, 0.5)$. It is of the form

$$JI(\beta) = [\text{Beta}(\beta, n\hat{p} + 1/2, n(1 - \hat{p}) + 1/2), 1] \quad (2.5.1)$$

where $\text{Beta}(\beta, a, b)$ is the $100(1 - \beta)\%$ quantile of the $\text{Beta}(a, b)$ distribution.

The Second Order Corrected interval is constructed based on the Edgeworth expansion to explicitly eliminate both the first and second order systematic bias in the coverage. Although the Edgeworth expansions are mainly regarded as asymptotic approximations, two-term Edgeworth expansions are very accurate for the two-sided problem, even for relatively small and moderate n . See Brown, Cai and DasGupta [7] for a detailed development of the two-sided confidence interval and discussions of accuracy coverage.

Cai [6] shows that these results are also true for the one-sided problem and that the second-order corrected intervals perform well for small and moderate sample sizes.

To define the Second Order Corrected Interval, let

$$\eta = \frac{1}{3}\kappa^2 + \frac{1}{6}; \quad \gamma_1 = -\left(\frac{13}{18}\kappa^2 + \frac{17}{18}\right); \quad \gamma_2 = \frac{1}{18}\kappa^2 + \frac{7}{36}; \quad \tilde{\mu} = \frac{n\hat{p} + \eta}{n + 2\eta}; \quad \text{and}$$

$V(\tilde{\mu}) = \tilde{\mu} - \tilde{\mu}^2$. Then the lower bound of the β -level one-sided confidence interval is defined by

$$CI_L = \tilde{\mu} - \kappa \sqrt{\frac{V(\tilde{\mu}) + \frac{\gamma_1 V(\tilde{\mu}) + \gamma_2}{n}}{n}}. \quad (2.5.2)$$

Using Monte Carlo simulations, for each generated sample, we compare the performance of confidence intervals based on Generalized Gamma Distribution with that of nonparametric confidence intervals, namely, the Jeffreys and the Second Order Corrected interval defined by (2.5.1) and (2.5.2), respectively.

2.6 Detection of Time Trend in Data

The techniques proposed in this dissertation are suitable only for data that is trend free. Thus, prior to application of the technique, data must be tested for existence of a time trend. There are several techniques that allow one to detect trend in data. The most

commonly used in reliability theory is the Laplace test, for which no compelling evidence has been produced to indicate that other tests have any advantages from the standpoint of power, computational ease or simplicity of interpretation of test results. Laplace's test is discussed in detail in Ascher and Feingold [18].

Laplace's test is based on the fact that under the identical distribution of the sample, the statistic

$$U = \frac{\frac{\sum_{i=1}^{n-1} x_i}{n-1} - \frac{x_n}{2}}{x_n \sqrt{\frac{1}{12(n-1)}}}$$

has approximately a standard normal distribution when n is relatively large. Here, x_i represents arrival times of errors measured from some defined "time-zero" point. Using tables of the tail probabilities for the standard normal distribution, we reject the level α hypothesis that the sample is identically distributed whenever $|U| > z_{\alpha/2}$ where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution $\Phi(z_{\alpha/2}) = 1 - \alpha/2$. Since very often one is interested in detecting only the trend in a particular direction, very often a one-sided hypothesis is tested and the trend is considered to be detected whenever $|U| > z_\alpha$ where $\Phi(z_\alpha) = 1 - \alpha$. The Laplace test is often used in reliability theory to detect the time dependency of failure data when the inter-arrival times are known, i.e. the inter-arrival times are the x_i . However, the NASA PRACA data discussed in Section 1, is not recorded in a manner which makes interarrival times available for analysis.

For data where error interarrival times are not readily available, we turn to the government standard MIL-HDBK-189, Reliability Growth Management, as suggested by Rigdon and Basu [34]. This handbook deals exclusively with reliability of repairable systems. The χ^2 goodness of fit test is one recommendation for test of trend in data. Suppose we partition the test time into k disjoint subintervals and observe the number N_i of failures in the i th interval which has length T_i . The test statistic is then

$$\begin{aligned} X^2 &= \sum_{i=1}^k \frac{(\text{observed \# of failures} - \text{expected \# of failures})^2}{\text{expected \# of failures}} \\ &= \sum_{i=1}^k \frac{(N_i - NP_i)^2}{NP_i} \end{aligned}$$

where $N = \sum_{i=1}^k N_i$ and $P_i = \frac{T_i}{t}$ with t representing the truncation time of the last sample.

Then the hypothesis of no trend is rejected for large values of X^2 . It has been the author's experience that many samples taken from the NASA PRACA data are trend free [36]. Once this fact has been confirmed, techniques for i.i.d samples may be applied.

2.7 Change Point Analysis

The problem of testing and estimating change points has attracted much attention in the literature since it originated in the field of quality control. Let X_1, X_2, \dots, X_n be a

sequence of observations where X_i has a distribution function $F_i(x)$. The change point problem refers to testing the existence of $k < n$ such that the first k distribution functions are identical and different from the last $n - k$ distribution functions which are also identical to each other. If the distributions F_1, F_2, \dots, F_n belong to a common parametric family $F(x | \theta)$, that is, $F_i(x) = F(x | \theta_i), i = 1, \dots, n$, the change point problem reduces to detecting whether there is a value $k, k < n$, such that

$$\alpha = \theta_1 = \theta_2 = \dots = \theta_k \neq \theta_{k+1} = \dots = \theta_n = \beta.$$

In the context of interval estimation, the existence of a change point may significantly change both the calculation of the confidence intervals and the inferences derived from them. When a random sample has a change point, k , it is no longer iid, therefore previous inferences that are based on the iid assumption are no longer valid and will introduce error into the confidence interval construction. Further, the question is whether the error is large enough to warrant detection and calculation of a change point when the sample size is small? These issues will be investigated in Section 3.

2.7.1 Formulation of the Problem

Since we are dealing with relatively small sample sizes, a maximum of one change point is a reasonable expectation while still providing sub-samples with sufficient size for computational purposes. The theory for a single change point can be easily extended to

the identification of multiple change points when sample size is sufficiently large to support such hypothesis.

To find a single change point, the hypothesis testing problem statement can be formulated as follows:

For the population parameters $\theta_i, i = 1, \dots, n$ the Null hypothesis is

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_n = \theta \text{ (unknown)}$$

versus the alternative hypothesis,

$$H_1 : \alpha = \theta_1 = \theta_2 = \dots = \theta_k \neq \theta_{k+1} = \dots = \theta_n = \beta$$

where $1 \leq k \leq n-1$ and is the parameter to be found.

2.7.2 Schwartz Information Criteria (SIC)

A non-Bayesian method for detecting the existence of a change point and concurrently estimating its value is suggested by Chen and Gupta [35] as the use of the Schwartz Information Criterion (SIC). This method will be used as a benchmark technique for simulations for the change point detection discussed in Section 4. The SIC method is a modification of the Akaike Information Criterion (AIC) for model selection in statistics (Akaike, 1973). This criterion has profoundly influenced developments in many areas of statistical analysis. The purpose of AIC is to choose which of the K available models is

the most appropriate for our data. The idea of AIC is as follows. When the number of parameters in the model grow, the pdf can fit the data more closely. However, this closeness is artificial and is likely to vanish when new data is obtained. Hence, it is necessary to guard against introducing too many parameters. The AIC assumes that if the model k is chosen, then an iid random sample of size n represents a random variable with some distribution $f_k(x | \Theta_k)$ where $\Theta_k = \{\theta_1, \theta_2, \dots, \theta_k\}$ and all components are non-zero.

Akaike proposed information criterion as a measure of model evaluation.

$AIC(k) = -2 \log L(\hat{\Theta}_k) + 2k$, $k = 1, 2, \dots, \mathbf{K}$, where $L(\theta)$ is a likelihood function and $\hat{\Theta}_k$ is the maximum likelihood estimator of Θ_k . A model which minimizes the $AIC(k)$ is considered to be the most appropriate model. Hence, Akaike suggests to penalize complexity of the model by adding $2k$ penalty to $-2 \log L(\hat{\Theta}_k)$.

Schwartz found that AIC criteria is not an asymptotically consistent estimator of model order, and modified the AIC to the SIC where:

$$SIC(k) = -2 \log L(\hat{\Theta}_k) + k \log n, \quad k = 1, 2, \dots, \mathbf{K},$$

where $k \log n$ is the penalty term. The SIC gives an asymptotically consistent estimate of the order of the true model.

To use the SIC for locating a change point, we use the principle of Information Criterion, to estimate \hat{k} such that $SIC(\hat{k})$ is minimal. To be specific, for the null hypothesis, H_0 ,

Chen and Gupta [35] found that $SIC(n) = n \log 2\pi + n \log \hat{\sigma}^2 + n + \log n$ where $\hat{\sigma}^2$ is the maximum likelihood estimator (MLE) of the variance of the distribution under H_0 . This estimator is:

$$\hat{\sigma}^2 = \frac{\hat{a}^{\frac{2}{\hat{\lambda}}}}{(\Gamma(\hat{b}))^2} \left(\Gamma(\hat{b}) \Gamma\left(\hat{b} + \frac{2}{\hat{\lambda}}\right) - \Gamma\left(\hat{b} + \frac{1}{\hat{\lambda}}\right)^2 \right)$$

by the Invariance Property of MLEs, where \hat{a} , \hat{b} and $\hat{\lambda}$ are the MLEs for the Generalized Gamma distribution based on a random sample of size n . These MLEs will be computed using the method suggested by Stacy & Mihram [37].

For the alternative hypothesis, H_1 , $SIC(k) = n \log 2\pi + k \log \hat{\sigma}_1^2 + (n - k) \log \hat{\sigma}_2^2 + n + \log n$ where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the MLE's of the variances of the two distributions separated by the change point. These estimators are obtained using the same method as for $\hat{\sigma}^2$, except the first k and the last $n - k$ observations, respectively, will be used. We accept H_0 if $SIC(n) < \min_{2 \leq k \leq n-2} SIC(k)$ and we accept H_1 if $SIC(n) > SIC(k)$ for some k . Then the estimate of the change point position, \hat{k} , is:

$$SIC(\hat{k}) = \min_{2 \leq k \leq n-2} SIC(k)$$

This non-Bayesian hypothesis testing method will be compared to the Bayesian hypothesis testing techniques discussed in Section 2.2 and developed further in Section 3.

3.0 TECHNICAL RESULTS

This chapter describes construction of Bayesian confidence intervals based on a set of i.i.d. observations $\underline{X} = \{x_i\}_{i=1}^n$ for the distribution function $F(z)$ and the inverse of the distribution function $F^{-1}(u)$, where z and u are known predetermined values. We assume that the sample is drawn from the Generalized Gamma distribution, discussed in Section 2.1, with the probability density function (pdf) given by (2.1.4). As previously discussed, this choice of pdf allows avoidance of a nonparametric set up which may lead to unreasonably wide confidence intervals when the sample size n is small. Yet the Generalized Gamma distribution provides flexibility to include many well-known distributions used in Reliability and Survival Analysis models (see Section 2.1).

Since no prior information on the values of parameters a , b and λ is available, the noninformative Jeffreys prior will be used for construction of Bayesian confidence intervals. It is shown that use of the Jeffreys prior leads to a proper posterior distribution and hence enables construction of these Bayesian confidence intervals for $F(z)$ and $F^{-1}(u)$. These confidence intervals will be studied with respect to their average length and coverage via Monte Carlo simulations and compared to the average length and coverage of the nonparametric confidence intervals described in Section 2.5.

This chapter also develops tests to determine if we have identically distributed sample values by exploring whether or not there is a change in the distribution of the data. These

tests are usually referred to as the change point analysis of the sample. In order to accomplish this goal, we shall analyze whether the sample, $\underline{X} = \{x_i\}_{i=1}^n$, is indeed i.i.d. or if for some $1 < k < n$, $\underline{X}_1 = \{x_i\}_{i=1}^k$ has a pdf $f_1(x)$ while $\underline{X}_2 = \{x_i\}_{i=k+1}^n$ has a pdf $f_2(x)$ where $f_1(x) \neq f_2(x)$. This is a change point detection and location. Taking the answer to this question into account, we shall then construct confidence bounds (1.1) and (1.2), and determine the impact of using confidence intervals that ignore a change point versus confidence intervals based on a detected change point. Monte Carlo simulations will explore precision of the suggested procedures for some sample models from the Generalized Gamma distribution family.

3.1 Derivation of the Jeffreys Prior

In this section, we derive the Jeffreys prior for the parameters a , b and λ , using equation (2.4.1). Since the sample $\underline{X} = \{x_i\}_{i=1}^n$ is i.i.d., by formula (2.1.4) the pdf of the sample is

$$f(\underline{X} | a, b, \lambda) = \prod_{k=1}^n f(x_k | a, b, \lambda) \text{ where}$$

$$f(x_i | a, b, \lambda) = \frac{\lambda}{\Gamma(b)a^b} x_i^{\lambda b - 1} e^{-\left(\frac{x_i^\lambda}{a}\right)}. \quad (3.1.1)$$

As discussed in Section 2.4, the Jeffreys prior defined by (2.4.1) is

$$\pi_J(\omega) = |\det H(\omega)|^{\frac{1}{2}}$$

where $\omega = \{\omega_k\}_{k=1}^3$ is a vector with components $\omega_1 = a$, $\omega_2 = b$, $\omega_3 = \lambda$.

Therefore, $H(\omega) = \{H_{i,j}\}$ is such that

$$H_{i,j} = -E \left(\frac{\partial^2}{\partial \omega_i \partial \omega_j} \log \prod_{k=1}^n f(x_k | \omega) \right) \quad i = 1, 2, 3 \quad j = 1, 2, 3. \quad (3.1.2)$$

Note that since $\underline{X} = \{x_i\}_{i=1}^n$ is an i.i.d. sample, the matrix, H , can be computed on the basis of only one observation, adding a factor of n to all elements of the matrix. In Bayesian inference, a constant factor based on n will appear in both the numerator and the denominator (see formula (2.2.2)). Hence, the constant will cancel. This factor can

therefore be ignored in future calculations of posterior distributions. We therefore perform all our calculations of Jeffreys prior on the basis of one sample value only. However, for completeness, the constant is shown in the final Jeffreys prior equation.

In order to derive the closed form expression for (3.1.2), differentiation, integration and matrix computation was performed using Mathematica version 5.1 software (Appendix 1). We apply the usual definition for the Gamma($\Gamma(x)$) function and define the Polygamma function $\Psi(m, y)$ as

$$\Psi(m, y) = \frac{d^{m+1}}{dy^{m+1}} \log \Gamma(y) \quad .$$

Hence, the elements of the symmetric matrix, H, are of the form:

$$H_{1,1} = \frac{1}{\lambda^2} \left(1 + \frac{\Gamma(1+b) \left((\log(a) + \Psi(0, 1+b))^2 + \Psi(1, 1+b) \right)}{\Gamma(b)} \right),$$

$$H_{1,2} = -\frac{b(\log(a) + \Psi(0, 1+b))}{a\lambda},$$

$$H_{1,3} = -\frac{\log(a) + \Psi(0, b)}{\lambda},$$

$$H_{2,2} = \frac{b}{a^2}, \quad H_{2,3} = \frac{1}{a},$$

$$H_{3,3} = \Psi(1, b).$$

Calculation of the square root of the determinant of this matrix yields the Jeffreys prior for the three parameters of the Generalized Gamma distribution for one observation. Incorporating sample independence, we obtain the Jeffreys prior for the three parameters of the Generalized Gamma distribution, based on a random sample $\underline{X} = \{x_i\}_{i=1}^n$ as

$$J\pi_{\underline{X}}(a, b, \lambda) = \sqrt{n} \frac{H(b)}{a \lambda} \quad \text{where} \quad H(b) = \sqrt{\Psi(1, b)(b^2 \Psi(1, b) - 1) - 1}. \quad (3.1.3)$$

It is easy to see that $J\pi_{\underline{X}}(a, b, \lambda)$ is decreasing as a and λ increase. From Figure 3, we see that $J\pi_{\underline{X}}(a, b, \lambda)$ is also decreasing rapidly as b increases. It is also clear that the value of $H(b)$ remains less than 1 unless b is infinitesimally small and close to zero.

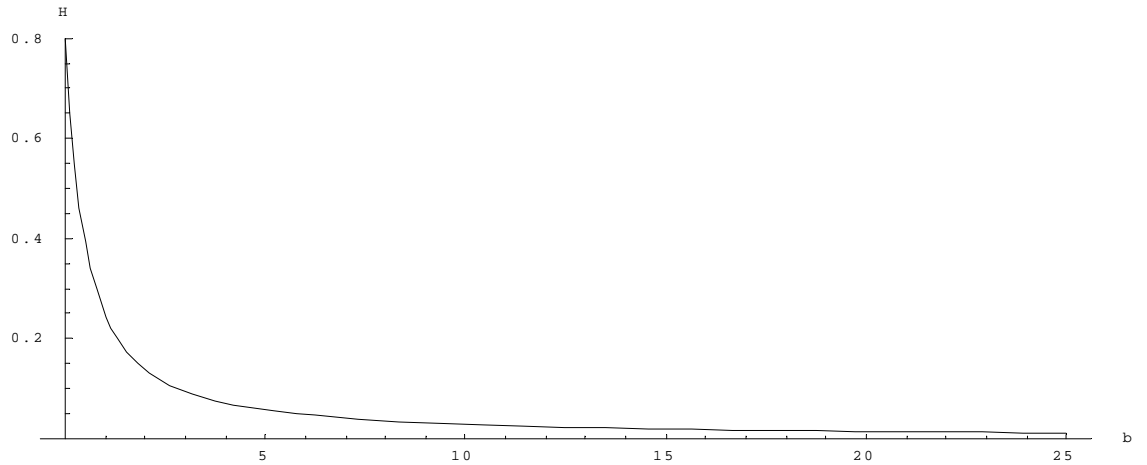


Figure 3: Plot of $H(b)$ vs. b

3.2 The Posterior Distribution of a, b, and λ

Having obtained the Jeffreys prior distribution, we now derive the posterior distribution of a, b and λ based on the given random sample. Using equation (2.2.3), where f_g is the pdf of the Generalized Gamma distribution given by formula (3.1.1) and $\pi_{\underline{X}}$ is the Jeffreys prior given by equation (3.1.3), the required posterior distribution of a, b and λ given the sample $\underline{X} = \{x_i\}_{i=1}^n$ is

$$\pi_p(a, b, \lambda | \underline{X}) \equiv \pi_p(a, b, \lambda) = \frac{\prod_{i=1}^n \frac{\lambda}{\Gamma(b) a^b} x_i^{\lambda b - 1} e^{-\left(x_i^\lambda / a\right)} \sqrt{n} \frac{H(b)}{a \lambda}}{\iiint \prod_{i=1}^n \frac{\lambda}{\Gamma(b) a^b} x_i^{\lambda b - 1} e^{-\left(x_i^\lambda / a\right)} \sqrt{n} \frac{H(b)}{a \lambda} da d\lambda db}.$$

Simplifying this expression, we obtain

$$\pi_p(a, b, \lambda) = \frac{\frac{H[b]}{\Gamma(b)^n} \frac{\lambda^{n-1}}{a^{bn+1}} P_n^{\lambda b - 1} \exp\left(-\frac{S_{n\lambda}}{a}\right)}{\iiint \frac{H[b]}{\Gamma(b)^n} \frac{\lambda^{n-1}}{a^{bn+1}} P_n^{\lambda b - 1} \exp\left(-\frac{S_{n\lambda}}{a}\right) da d\lambda db} \quad (3.2.1)$$

where $P_n = \prod_{i=1}^n x_i$, (3.2.2)

and $S_{n\lambda} = \sum_{i=1}^n x_i^\lambda$. (3.2.3)

Note that the denominator in (3.2.1) is independent of the parameters a, b and λ . Thus the denominator can be replaced by a normalizing constant $C(\underline{X})$, a function of the random sample \underline{X} and its size n. As will be shown in Section 3.5, if $a, b \in (0, \infty)$ and

$\lambda \in (0, \lambda_0)$, then the posterior (3.2.1) is proper and the constant $C(\underline{X})$ is finite and equal to a finite value $\frac{1}{K}$.

Thus the posterior distribution for a , b and λ as shown in equation (3.2.1) can be written as

$$\pi_p(a, b, \lambda) = K \frac{H[b]}{\Gamma(b)^n} \frac{\lambda^{n-1}}{a^{bn+1}} P_n^{\lambda b-1} \exp\left(-\frac{S_{n\lambda}}{a}\right) \quad (3.2.4)$$

where K is a finite normalizing constant.

3.3 The Posterior Distribution of $F(z)$ and $F^{-1}(u)$

To construct a confidence interval, define $F(z)$, where $z \in \mathfrak{R}$ is a fixed value and F is the cumulative distribution function of the Generalized Gamma distribution (3.1.1). Then $F(z)$ is a function of a , b , and λ , and hence a new random variable itself under the Bayesian paradigm. Transformation of variables in (3.2.2) and subsequent integration will then provide a pdf of the new random variable $F(z)$.

We begin by introducing the Regularized Gamma function

$$Q[a, t] = \frac{\Gamma[a, t]}{\Gamma[a]} = \frac{1}{\Gamma[a]} \int_t^\infty s^{a-1} e^{-s} ds = \xi$$

and the Inverse Regularized Gamma function, $Q^{-1}[a, t]$ such that $Q^{-1}[a, \xi] = t$.

Then the cdf of the Generalized Gamma distribution at the point $z \in \mathfrak{R}$ can be written as

$$F(z) = \int_0^z f(x | a, b, \lambda) d\lambda = 1 - \frac{\Gamma[b, \frac{z^\lambda}{a}]}{\Gamma[b]} = 1 - Q[b, \frac{z^\lambda}{a}]. \quad (3.3.1)$$

For a fixed $z \in \mathfrak{R}$, define the new random variable $c_z = F(z)$ and a new parameter

$w = \frac{z^\lambda}{a}$. Then

$$c_z = 1 - Q[b, w] \quad \text{and} \quad w = Q^{-1}[b, 1 - c_z]. \quad (3.3.2)$$

Note that if random variables X and Y are related as $X=h(y)$, then the pdf $g(y)$ of Y can be obtained from the pdf $p(x)$ of X using the common transformation formula

$$p(x) = g(h(y))h'(y), g(y) = p(h^{-1}(x))(h^{-1}(x))'. \quad (3.3.3)$$

Going back to equation (3.2.4) and applying the transformation formula (3.3.3) for

$w = \frac{z^\lambda}{a} \Rightarrow a = \frac{z^\lambda}{w}$, we derive that the joint pdf of w , b and λ is

$$\pi_p(w, b, \lambda) = \pi_p\left(\frac{z^\lambda}{w}, b, \lambda\right) \left| \frac{\partial a}{\partial w} \right| = K \frac{H[b]}{\Gamma(b)^n} \frac{\lambda^n}{\lambda w} \left(\frac{w}{z^\lambda} \right)^{bn} P_n^{\lambda b-1} \exp\left(-\frac{w}{z^\lambda} S_{n\lambda}\right).$$

Choosing w defined by (3.3.2), we derive

$$\begin{aligned} \pi_p(c_z, b, \lambda) &= \pi_p(w = Q^{-1}[b, 1 - c_z], b, \lambda) \left| \frac{\partial w}{\partial c_z} \right| = \\ &= K H[b] \left(\frac{P_n}{z^n} \right)^{\lambda b} \left(\frac{\lambda Q^{-1}[b, 1 - c_z]^b}{\Gamma(b)} \right)^{n-1} \exp\left(Q^{-1}[b, 1 - c_z] \left(1 - \frac{S_{n\lambda}}{z^\lambda} \right) \right). \end{aligned}$$

Integrating this equation with respect to b and λ , we arrive at the posterior pdf of the new random variable c_z

$$\begin{aligned} \pi_p(c_z) &= \\ K \iint_{b, \lambda} H[b] \left(\frac{P_n}{z^n} \right)^{\lambda b} \left(\frac{\lambda Q^{-1}[b, 1 - c_z]^b}{\Gamma(b)} \right)^{n-1} \exp\left(Q^{-1}[b, 1 - c_z] \left(1 - \frac{S_{n\lambda}}{z^\lambda} \right) \right) db d\lambda. \end{aligned} \quad (3.3.4)$$

Similar to this construction of the pdf of $c_z = F(z)$ we shall derive the pdf of $F^{-1}(u)$, for a fixed $u \in [0, 1]$, where F^{-1} is the inverse of F , the cumulative distribution function of the Generalized Gamma distribution (3.1.1) as defined by (3.3.1).

For a fixed $u \in [0,1]$, define a new random variable $\rho_u = F^{-1}(u) \Rightarrow F(\rho_u) = u$. Then

from (3.3.1), $u = 1 - Q[b, \frac{\rho_u^\lambda}{a}]$ so that

$$a = \frac{\rho_u^\lambda}{Q^{-1}[b, 1-u]} .$$

Using the joint pdf of a, b and λ (3.2.4) and applying the transformation formula (3.3.3)

$$\text{for } a = \frac{\rho_u^\lambda}{Q^{-1}[b, 1-u]} ,$$

we derive

$$\pi_p(\rho_u, b, \lambda) = \pi_p\left(\frac{\rho_u^\lambda}{Q^{-1}[b, 1-u]}, b, \lambda\right) |\det J|$$

where J is the matrix representing the Jacobian of the transformation and has the form

$$J = (J_{ij}) , \quad i, j = 1, 2, 3 , \text{ where}$$

$$J_{ij} = \frac{\partial \alpha_i}{\partial \beta_j} \text{ and } \alpha_1 = a , \beta_1 = \rho_u , \alpha_2 = \beta_2 = b , \alpha_3 = \beta_3 = \lambda .$$

Performing the calculation provides

$$\pi_p(\rho_u, b, \lambda) = K \frac{H[b]}{\rho_u} \left(\frac{P_n}{\rho_u^n} \right)^{\lambda b} \left(\frac{\lambda Q^{-1}[b, 1-u]^b}{\Gamma(b)} \right)^n \exp \left(- \frac{Q^{-1}[b, 1-u]}{\rho_u^\lambda} S_{n\lambda} \right) . \quad (3.3.5)$$

Thus by integrating (3.3.5) with respect to b and λ we arrive at the posterior distribution (pdf) for the new random variable ρ_u

$$\pi_p(\rho_u) = K \iint_{b,\lambda} \frac{H[b]}{\rho_u} \left(\frac{P_n}{\rho_u^n} \right)^{\lambda b} \left(\frac{\lambda Q^{-1}[b, 1-u]^b}{\Gamma(b)} \right)^n \exp \left(-\frac{Q^{-1}[b, 1-u]}{\rho_u^\lambda} S_{n\lambda} \right) db d\lambda \quad . \quad (3.3.6)$$

For both (3.3.4) and (3.3.6), K is the constant which ensures that integration of $\pi_p(c_z)$ over the domain of $c_z \in [0,1]$ and the integration of $\pi_p(\rho_u)$ over the domain of $\rho_u \in [0, \infty]$ yields unity.

3.4 Confidence Intervals

Having derived the posterior pdfs for $c_z = F(z)$ and $\rho_u = F^{-1}(u)$, we are now ready to construct confidence intervals for these random variables.

Given some value of $\beta \in [0,1]$, equation (3.3.4) for $\pi_p(c_z)$ can be used to find a value U such that

$$P(c_z \geq U) = \int_U^1 \pi_p(c_z) dc_z = \beta \quad \text{where } c_z = F(z), \quad z \in \mathfrak{R}. \quad (3.3.7)$$

Thus, the first problem (1.1) stated in Chapter 1 is resolved. The interval $[U,1]$ is the $\beta*100\%$ confidence interval for $c_z = F(z)$.

Similarly, given some value of $\beta \in [0,1]$, equation (3.3.6) for $\pi_p(\rho_u)$ can be used to find Z such that

$$P(\rho_u \leq Z) = \int_0^z \pi_p(\rho_u) d\rho_u = \beta \quad \text{where } \rho_u = F^{-1}(u), u \in [0,1]. \quad (3.3.8)$$

Thus, for fixed $u \in [0,1]$ and β , we have the solution for Z such that

$$P(\rho_u \leq Z) = P(F^{-1}(u) \leq Z) = P(F(Z) \geq u) = \beta$$

which is the solution of the problem (1.2) stated in Chapter 1. The interval $(0, Z)$ is the $\beta*100\%$ confidence interval for $\rho_u = F^{-1}(u)$.

Note that construction of the confidence intervals above is feasible only if the pdfs $\pi_p(c_z)$ and $\pi_p(\rho_u)$ are proper densities (i.e. integrate to unity). The latter can be achieved provided the pdf (3.2.4) is integrable in $a, b, \lambda \in (0, \infty)$. Justification of this matter is the topic of the next section.

3.5 Justification of the Approach

As previously discussed, the theory above is applicable only if the posterior distribution defined by (3.2.1) is proper for $a, b \in (0, \infty)$ and $\lambda \in (0, \lambda_0)$. In order to prove this, we must show that the denominator in (3.2.1) is finite.

First, let us consider asymptotic behavior of $H(b)$ as $b \rightarrow 0^+$ and $b \rightarrow \infty$. Denote $H_1(b) = H^2(b)$ and observe that by Abramovitz & Stegun [38] the polygamma function,

$\Psi(1, b) = \sum_{k=0}^{\infty} (b+k)^{-2}$ for $b \neq 0, 1, 2, \dots$. Hence, by direct calculations, we obtain that

$$H_1(b) = \left(1 + \sum_{l=1}^{\infty} \frac{b^2}{(b+l)^2} \right) \sum_{k=1}^{\infty} \frac{1}{(b+k)^2} - 1,$$

so that, by formulae 0.233.3 and 9.71 of reference [39],

$$\lim_{b \rightarrow 0^+} H_1(b) = \sum_{k=1}^{\infty} k^{-2} - 1 = \frac{\pi^2}{6} - 1.$$

In order to derive asymptotic expression for $H(b)$ as $b \rightarrow \infty$, note that by Abramovitz & Stegun [38], $\Psi(1, b) \approx b^{-1} + (2b^2)^{-1} + (6b^3)^{-1} + \dots$ as $b \rightarrow \infty$, thus, $H_1(b) \approx (4b^2)^{-1}$.

Summarizing, we obtain

$$\lim_{b \rightarrow 0^+} H(b) = \sqrt{\frac{\pi^2}{6} - 1} \approx 0.803 < \infty, \quad \lim_{b \rightarrow \infty} bH(b) = \frac{1}{2} \quad (3.5.1)$$

In order to prove the denominator of (3.2.1) is finite, first integrate the denominator with respect to a , deriving

$$p(b, \lambda | \underline{X}) = CH(b)\Gamma(bn)[\Gamma(b)]^{-n} P_n^{\lambda b-1} S_{n\lambda}^{-bn} \lambda^{n-1} \quad (3.5.2)$$

In what follows we shall need asymptotics of the function $G_n(b) = \Gamma(bn)[\Gamma(b)]^{-n} n^{-nb}$ for fixed values of n as $b \rightarrow 0^+$ and $b \rightarrow \infty$. If $b \rightarrow 0^+$, using exact representation 8.334 from reference [39], we obtain $G_n(b) = n^{-nb} \pi^{1-n} [\sin(\pi b)]^n \Gamma(1-b) / [\Gamma(1-nb) \sin(\pi nb)]$, so that

$$G_n(b) = \Gamma(bn)[\Gamma(b)]^{-n} n^{-nb} \sim n^{-1} b^{n-1} \text{ as } b \rightarrow 0^+. \quad (3.5.3)$$

If $b \rightarrow \infty$, then formula 8.327 from reference [39] yields

$$G_n(b) \sim b^{(n-1/2)} (2\pi)^{(1-n/2)} n^{-1/2} \text{ as } b \rightarrow \infty. \quad (3.5.4)$$

Now, let us denote $V_{n\lambda} = P_n^\lambda S_{n\lambda}^{-n} n^n$. Note that since the geometric mean never exceeds the arithmetic mean, $V_{n\lambda} \leq 1$ and equality is attained only for identical sample values X_i .

Denote $u_{n\lambda} = \ln \left(\frac{1}{V_{n\lambda}} \right)$ and observe that, hence, $u_{n\lambda} > 0$.

Recall that $p(b, \lambda | \underline{X}) = CH(b)G_n(b)V_{n\lambda}^b \lambda^{n-1}$. Using (3.5.1), (3.5.3) and (3.5.4), we derive that

$$p(b, \lambda | \underline{X}) \sim \begin{cases} C_{n1} \lambda^{n-1} b^{n-1} \exp(-bu_{n\lambda}), & b \rightarrow 0^+ \\ C_{n2} \lambda^{n-1} b^{(n-3)/2} \exp(-bu_{n\lambda}), & b \rightarrow \infty \end{cases} \quad (3.5.5)$$

for some positive constants C_{n1} and C_{n2} independent of b and λ . Hence, integration of $p(b, \lambda | \underline{X})$ with respect to $b \in (0, \infty)$ yields

$$p(\lambda | \underline{X}) = C_n \lambda^{n-1} \left[u_{n\lambda}^{-n} \gamma(n, u_{n\lambda}) + u_{n\lambda}^{\left(\frac{1-n}{2}\right)} \Gamma\left(\frac{n-1}{2}, u_{n\lambda}\right) \right] \quad (3.5.6)$$

where $\gamma(n, u_{n\lambda})$ and $\Gamma\left(\frac{n-1}{2}, u_{n\lambda}\right)$ are incomplete gamma functions defined by 8.350

from reference [39], and positive constant C_n is independent of λ .

Denote

$$\kappa_n = n^{-1} \sum_{i=1}^n (\ln X_i)^2 - \left(n^{-1} \sum_{i=1}^n (\ln X_i) \right)^2 > 0, \quad \rho_n = \ln \left(\max_i X_i \right) - n^{-1} \sum_{i=1}^n (\ln X_i) > 0. \quad (3.5.7)$$

In order to assess integrability of $p(\lambda | \underline{X})$ we need to study asymptotic behavior

of $u_{n\lambda}$ as $\lambda \rightarrow 0^+$ and $\lambda \rightarrow \infty$. Note that as $\lambda \rightarrow \infty$, the value of $S_{n\lambda}$ is dominated by

its largest term, i.e. $S_{n\lambda} \sim (\max_i X_i)^\lambda$. If $\lambda \rightarrow 0^+$, then

$$X_i^\lambda = \exp(\lambda \ln X_i) \sim 1 + \lambda \ln X_i + 0.5\lambda^2 (\ln X_i)^2 + \dots,$$

so that

$$n^{-1} S_{n\lambda} \sim 1 + \lambda n^{-1} \sum_{i=1}^n \ln X_i + 0.5\lambda^2 n^{-1} \sum_{i=1}^n (\ln X_i)^2.$$

Calculation of $u_{n\lambda}$ in view of the above and asymptotic formula

$\ln(1+x) = x - 0.5x^2 + o(x^2)$ as $x \rightarrow 0$, results in the following asymptotic expression for

$u_{n\lambda}$

$$u_{n\lambda} = n \ln(n^{-1} S_{n\lambda}) - \lambda \ln P_n \sim \begin{cases} 0.5n\kappa_n \lambda^2, & \lambda \rightarrow 0^+ \\ n\rho_n \lambda, & \lambda \rightarrow \infty \end{cases} \quad (3.5.8)$$

where κ_n and ρ_n are defined in (3.5.7).

Now, taking into account that $\gamma(n, u_{n\lambda}) \sim n^{-1} x^n$ as $x \rightarrow 0^+$ and $\Gamma\left(\frac{n-1}{2}, x\right) \sim x^{(n-3)/2} e^{-x}$ as

$x \rightarrow \infty$ (see 8.354.1 and 8.357 of from reference [39]), we arrive at

$$p(\lambda | \underline{X}) \sim \begin{cases} C_{n3} \lambda^{n-1} + C_{n4}, & \lambda \rightarrow 0^+ \\ C_{n5} \lambda^{-1} + C_{n6} \lambda^{n-2} \exp(-n\rho_n \lambda), & \lambda \rightarrow \infty \end{cases}$$

where C_{ni} , $i=3,4,5,6$ are independent of λ . Hence, the posterior distribution in (3.2.1) is a proper density whenever the domain of λ is bounded from above.

3.6 Change Point Detection and Location

Thus far, we have assumed that random samples represent a single Generalized Gamma distribution and generated confidence intervals for the distribution function, $F(z)$ and quantiles, $\rho_u = F^{-1}(u)$. We now turn our attention towards detecting if a single change point exists in the random sample, implying that a random sample represents two distinct

Generalized Gamma distributions. The latter will change calculation of the interval estimators.

We return to equation (2.1.4) to represent the Generalized Gamma distribution $f_g(x|a, b, \lambda)$ for a random sample $\underline{X} = \{x_i\}_{i=1}^n$. When k is the change point to be detected, the hypothesis testing problem, as described in Section 2.7, can be formally restated as: testing null hypothesis $H_0 : k = n$ versus the alternative hypothesis, $H_1 : k \neq n$, $1 \leq k \leq n-1$.

The change point k denotes the position in the random sample where change in distribution occurs. Hence, $\underline{X}_k = \{x_i\}_{i=1}^k$ represents one generalized gamma distribution and $\underline{X}_{n-k} = \{x_i\}_{i=k+1}^n$ represents a different generalized gamma distribution. We therefore have two problems to solve. The first one is to test the hypothesis that there is a change point in the sample. The second one is to estimate the location of the change point provided it has been detected. Later, all techniques developed in this section will be compared by Monte Carlo simulations to the Schwartz Information Criterion (SIC) which provides both change point detection and location computations.

3.6.1 Problem Formulation for Change Point Detection and Location

For a Generalized Gamma distribution without a change point, the pdf of the sample

$\underline{X} = \{x_i\}_{i=1}^n$ is of the form

$$f_g(\underline{X} | \theta, k = n) = \frac{\lambda^n}{(\Gamma(b))^n a^{bn}} \left(\prod_{i=1}^n x_i \right)^{\lambda b - 1} e^{-\sum_{i=1}^n x_i^\lambda / a}, \quad \theta = (a, b, \lambda).$$

For a Generalized Gamma distribution with a change point k , $k < n$, the pdf of the sample is of the form

$$f_g(\underline{X} | \theta_1, \theta_2, k) = \frac{\lambda_1^k \lambda_2^{n-k}}{(\Gamma(b_1))^k (\Gamma(b_2))^{n-k} a_1^{b_1 k} a_2^{b_2 (n-k)}} \left(\prod_{i=1}^k x_i \right)^{\lambda_1 b_1 - 1} \left(\prod_{i=k+1}^n x_i \right)^{\lambda_2 b_2 - 1} e^{-\sum_{i=1}^k x_i^{\lambda_1} / a_1 - \sum_{i=k+1}^n x_i^{\lambda_2} / a_2}.$$

We are going to test hypothesis H_0 using Bayes testing procedure with Jeffreys noninformative prior developed in Section 3.1. Recall that Jeffreys noninformative prior for a random sample $\underline{X} = \{x_i\}_{i=1}^n$ from the Generalized Gamma distribution is given by equation (3.1.3):

$$J\pi_{\underline{X}}(a, b, \lambda) = \sqrt{n} \frac{H(b)}{a \lambda} \quad \text{where} \quad H(b) = \sqrt{\Psi(1, b)(b^2 \Psi(1, b) - 1) - 1} \quad \text{and}$$

$$\Psi(1, b) = \frac{d^2}{db^2} \log \Gamma(b) .$$

Then the marginal distribution of the random sample under the null hypothesis H_0 is of the form:

$$\begin{aligned}
m^0(\underline{X} | k = n) &= \int_{\theta} f_g(\underline{X} | \theta, n) J \pi_{\underline{X}}(\theta) d\theta \\
&= \sqrt{n} \iiint_{a,b,\lambda} f_g(\underline{X} | a, b, \lambda) \frac{H(b)}{a \lambda} da db d\lambda = \sqrt{n} \iint_{b,\lambda} H(b) \Gamma(bn) [\Gamma(b)]^{-n} P_n^{\lambda b-1} S_{n\lambda}^{-bn} \lambda^{n-1} db d\lambda \\
&= \sqrt{n} I_n(\underline{X})
\end{aligned}$$

$$\text{where } P_n = \prod_{i=1}^n x_i \text{ and } S_{n\lambda} = \sum_{i=1}^n x_i^{\lambda} \text{ and } I_n(\underline{X}) = \iint_{b,\lambda} H(b) \Gamma(bn) [\Gamma(b)]^{-n} P_n^{\lambda b-1} S_{n\lambda}^{-bn} \lambda^{n-1} db d\lambda.$$

The marginal distribution of the random sample under the alternative hypothesis, H_1 , is of the form:

$$\begin{aligned}
m^A(\underline{X} | k) &= \iint_{\theta_1, \theta_2} f_g(\underline{X} | \theta_1, \theta_2, k) J \pi_{\underline{X}}(\theta_1) J \pi_{\underline{X}}(\theta_2) d\theta_1 d\theta_2 \\
&\propto \int \dots \int f_g(\underline{X} | a_1, b_1, \lambda_1, a_2, b_2, \lambda_2) \frac{H(b_1) H(b_2)}{a_1 a_2 \lambda_1 \lambda_2} da_1 da_2 db_1 db_2 d\lambda_1 d\lambda_2 \\
&= \sqrt{n-k} \iiint_{a_2, b_2, \lambda_2} f_g(\underline{X}_{k+1}^n | a_2, b_2, \lambda_2, n-k) \frac{H(b_2)}{a_2 \lambda_2} \left(\sqrt{k} \iiint_{a_1, b_1, \lambda_1} f_g(\underline{X}_1^k | a_1, b_1, \lambda_1, k) \frac{H(b_1)}{a_1 \lambda_1} da_1 db_1 d\lambda_1 \right) da_2 db_2 d\lambda_2 \\
&= \sqrt{k(n-k)} I_{n-k}(\underline{X}_{k+1}^n) I_k(\underline{X}_1^k)
\end{aligned}$$

where $1 \leq k \leq n-1$, $\underline{X}_1^k = \{x_i\}_{i=1}^k$ and $\underline{X}_{k+1}^n = \{x_i\}_{i=k+1}^n$,

$$I_k(\underline{X}_1^k) = \iint_{b_1, \lambda_1} H(b_1) \Gamma(b_1 k) [\Gamma(b_1)]^{-k} P_k^{\lambda_1 b_1-1} S_{k\lambda_1}^{-b_1 k} \lambda_1^{k-1} db_1 d\lambda_1 \text{ and}$$

$$I_{n-k}(\underline{X}_{k+1}^n) = \iint_{b_2, \lambda_2} H(b_2) \Gamma(b_2(n-k)) [\Gamma(b_2)]^{-(n-k)} P_{(n-k)}^{\lambda_2 b_2-1} S_{(n-k)\lambda_2}^{-b_2(n-k)} \lambda_2^{(n-k)-1} db_2 d\lambda_2.$$

$$P_{n-k} = \prod_{i=k+1}^n x_i \text{ and } S_{(n-k)\lambda} = \sum_{i=k+1}^n x_i^{\lambda}.$$

For convenience, in what follows we shall often drop \underline{X} and denote $m^0(\underline{X}|n)$ as m^0 and $m^A(\underline{X}|k)$ as m_k^A . Hence, using the Jeffreys prior, we obtain marginal distributions of the sample \underline{X} , m^0 and m_k^A , under hypotheses H_0 and H_1 , respectively.

In order to perform Bayesian analysis, we assign a prior distribution to k . We shall use the prior distribution suggested by Chen and Gupta[35]

$$g(k|p) = \begin{cases} p, & k = n \\ \frac{1-p}{n-1}, & k \neq n \end{cases} \quad \text{where } p \in [0,1] \quad 1 \leq k \leq n. \quad (3.6.1)$$

Then the joint probability of \underline{X} and H_0 given p is

$P(\underline{X}, H_0 | p) = P(\underline{X} | k = n)g(k|p) = m^0 g(n) = m^0 p$, and the joint probability of \underline{X} and

H_1 given p is $P(\underline{X}, H_1 | p) = P(\underline{X} | k \neq n)g(k \neq n | p) = \sum_{k=1}^{n-1} m_k^A g(k) = (1-p)M_n$, where

$$M_n = \frac{\sum_{j=1}^{n-1} m_j^A}{n-1}. \quad \text{The two equations imply that}$$

$$P(\underline{X} | p) = P(\underline{X}, H_0 | p) + P(\underline{X}, H_1 | p) = m^0 p + (1-p)M_n. \quad (3.6.2)$$

The ratio of the probabilities of H_0 and H_1 given \underline{X} and p is given by the Bayes Factor

$$BF = \frac{P(\underline{X}, H_0 | p)}{P(\underline{X}, H_1 | p)} = \frac{pm^0}{(1-p)M_n}. \quad (3.6.3)$$

The posterior distribution of k can be computed from Bayes Rule:

$$h(k|p) = P(k | \underline{X}, p) = \frac{P(k, \underline{X} | p)}{P(\underline{X} | p)} = \frac{P(\underline{X} | k)g(k|p)}{P(\underline{X} | p)}.$$

Thus,

$$h(k = n | p) = P(k = n | \underline{X}, p) = \frac{m^0 p}{P(\underline{X} | p)}, \quad \text{and}$$

$$h(k \neq n) = P(k \neq n | \underline{X}) = \frac{(1-p)m_k^A}{(n-1)P(\underline{X} | p)}.$$

Combination of the latter with (3.6.2) provides the posterior distribution of the change point as:

$$h(k | p) = \begin{cases} \frac{m^0 p}{m^0 p + (1-p)M_n} & k = n \\ \frac{\left(\frac{1-p}{n-1}\right)m_k^A}{m^0 p + (1-p)M_n} & k \neq n \end{cases} \quad (3.6.4)$$

Valid estimators for the location of change point \hat{k} are the mean, median or mode of the distribution defined by (3.6.4).

The mean of this distribution is equal to

$$k_{mean} = \frac{1}{m^0 p + (1-p)M_n} \left(nm^0 p + \frac{1-p}{n-1} \sum_{k=1}^{n-1} km_k^A \right) = \frac{n(n-1)m^0 p + (1-p)M_E}{(n-1)m^0 p + (1-p)M}, \quad (3.6.5)$$

$$\text{where } M = \sum_{k=1}^{n-1} m_k^A \text{ and } M_E = \sum_{k=1}^{n-1} km_k^A.$$

Before calculating the median, recall that the Bayesian test detects a change point

$$\text{whenever } BF = \frac{pm^0}{(1-p)M_n} < 1 \text{ is true, which is equivalent to } \frac{m^0 p}{m^0 p + (1-p)M_n} < .5.$$

When $\frac{m^0 p}{m^0 p + (1-p)M_n} < .5$, the median of the distribution (3.6.4) is the solution

$k_{med} \leq n-1$ of the equation

$$\frac{1-p}{(n-1)(m^0 p + (1-p)M_n)} \sum_{j=1}^{k_{med}} m_j^A = .5 \quad (3.6.6)$$

The mode of the distribution (3.6.4) is k_{mode} such that $h(k_{mode}) = \max_{1 \leq k \leq n} h(k)$. Thus,

$$\begin{aligned} k_{mode} &= \arg \max_{1 \leq k \leq n} \left(\frac{m^0 p}{m^0 p + (1-p)M_n}, \frac{\left(\frac{1-p}{n-1}\right) m_k^A}{m^0 p + (1-p)M_n} \right) \\ &= \arg \max_{1 \leq k \leq n} \left(m^0 p, \left(\frac{1-p}{n-1}\right) m_k^A \right) \end{aligned} \quad (3.6.7)$$

Note that this computation can yield $k_{mode} = n$, in spite of Bayes factor pointing out that the change point in the sample exists, thereby rejecting H_1 . Thus the computation of k_{mode} can also be used as a change point detection technique.

3.6.2 Bayes Factor Change Point Detection and Location

Using the Bayesian model described in Section 3.6.1, the Bayes Factor may be used for

change point detection. When $BF = \frac{pm^0}{(1-p)M_n} < 1$, we reject H_0 and proceed to estimate

the value of \hat{k} as defined by equations (3.6.5), (3.6.6) or (3.6.7). Monte Carlo simulations in Section 4 will provide all three estimators in order to compare estimator accuracy.

Here, p , as the probability of no change point, is a value known or assigned in advance.

When p is unknown, there are three alternatives:

1. Assign a value that supports a-priori assumptions about the change point. For example, letting $p = \frac{1}{n}$ assigns equal prior probability for any change point location. Or, letting $p = .5$ assigns equal prior probability to the null and alternative hypothesis. For $p = \frac{1}{n}$, the Bayes Factor becomes:

$$BF = \frac{m^0}{M} \quad \text{where} \quad M = \sum_{j=1}^{n-1} m_j^A. \quad (3.6.8)$$

And for $p = .5$, the Bayes Factor becomes:

$$BF = \frac{(n-1)m^0}{M} \quad \text{where} \quad M = \sum_{j=1}^{n-1} m_j^A. \quad (3.6.9)$$

2. Maximize $P(\underline{X} | p)$ with respect to p . Using equation (3.6.2), we would maximize: $P(\underline{X} | p) = m^0 p + (1-p)M_n$. When $m^0 = M$, this probability is 1, thus a value of p maximizing this expression does not exist. Otherwise, the maximizing p occurs at endpoints, $p = 0$ (implying rejection of H_0) when

$m^0 < M$ and $p = 1$ (implying rejection of H_1) when $m^0 > M$. This duplicates the result of setting $p = \frac{1}{n}$.

3. Assign a prior distribution to p , $\pi(p|\underline{\phi})$ and use equation (2.2.6) to derive the Bayes Factor value, which will be a function of $\underline{\phi}$. The prior distribution must be such that

- it is defined only for $p \in [0,1]$,
- provides a maximum for some $\underline{\phi}$ value using equation (2.2.7).

We tried to use the Beta distribution, $Beta(p|\alpha, \beta)$, as a prior distribution for p , however, expression (2.2.7) is maximized by either $p=0$ or $p=1$. Since there are no other distributions commonly used for a random variable $p \in [0,1]$, we are not going to implement hierarchical Bayes method in this dissertation.

Thus, we shall only study Bayes Factor based testing techniques for a fixed value of p .

In our simulations, we used both $p = \frac{1}{n}$ and $p = .5$. The corresponding change point

location estimators are obtained by substituting the specified value of p into equations (3.6.5), (3.6.6) and (3.6.7). The resulting formulae are shown in Table 1. Section 4

provides results of Monte Carlo simulations based on Table 1. It also studies the accuracy of the estimators of change point location in the case where the change point is detected.

3.6.3 Alternative Bayes Technique for Change Point Detection and Location

Alternatively, the Bayesian technique of assigning a prior distribution for k can use only the posterior distribution of k , ignoring the Bayes Factor, for change point detection. This section provides three alternative techniques for detection and estimation of location of the change point. Based on the results of Section 3.6.2, we only use values

$p = \frac{1}{n}$ and $p = .5$ when a value of p is required.

Recall that for the posterior distribution mode calculation in (3.6.7), the result can yield $k_{mode} = n$, independent of the decision delivered by the Bayes Factor. One can use $\hat{k} = k_{mode}$ for detection and estimation of location of the change point, deciding to reject H_1 whenever $k_{mode} = n$, and to reject H_0 whenever $k_{mode} \neq n$.

Recall that

$$k_{mode} = \arg \max_{1 \leq k \leq n} \left(m^0 p, \left(\frac{1-p}{n-1} \right) m_k^A \right) \neq n \Leftrightarrow m^0 p < \left(\frac{1-p}{n-1} \right) m_k^A \quad \text{for some } k \quad (3.6.10)$$

However, a value of p is still required for these calculations. In what follows, just the

same as in the case of Bayes factor calculation, we shall use $p = \frac{1}{n}$ and $p = .5$

Alternatively, note that from Section 3.6.1 we have m^0 and m_k^A , the distribution of the random sample given a specific change point k , derived on the basis of Bayesian techniques. For hypothesis testing one can use a statistic which is an analog of the likelihood ratio test statistic:

$$\omega(\underline{X}) = \frac{L(n | \underline{X})}{\sup_{1 \leq k \leq n-1} L(k | \underline{X})}$$

where $L(k | \underline{X}) = \prod_{i=1}^n m(x_i | k) = m(\underline{X} | k)$, the marginal distribution of the random sample, given a change point, k or no change point ($k = n$). Then the test statistic can be written as:

$$\omega(\underline{X}) = \frac{m^0}{\max_{1 \leq k \leq n-1} m_k^A} \quad (3.6.12)$$

Note that test statistic (3.6.12) is independent of the value of p . The null hypothesis, H_0 , is accepted if $\omega(\underline{X}) \geq A$ and rejected if $\omega(\underline{X}) < A$ where $A > 0$ is a threshold which is specified in advance. As a general rule, $A=1$. One sets $A < 1$ only if stronger evidence of the alternative hypothesis is required for acceptance, and $A > 1$ if stronger evidence of the null hypothesis is required for acceptance. Then, if this test detected a change point, we have:

$$\max_{1 \leq k \leq n-1} m_k^A > m^0 \quad (3.6.13)$$

and the estimator of the location of change point is $\hat{k} = k_{mode}$.

3.6.4 Summary

Table 1 summarizes the tests suggested in Section 3.6, and the estimators of change

point location based on a random sample of size n . We denote $M = \sum_{j=1}^{n-1} m_j^A$, $M_n = \frac{M}{n-1}$

and $M_E = \sum_{j=1}^{n-1} j m_j^A$. Section 4 will provide Monte Carlo simulation study of the tests and

estimators in Table 1.

Table 1: Change Point Detection/Estimation Techniques

Method	Hypothesis Test (Reject H_0)	P Estimator	k Estimator (rounded to integer)
Schwartz Information Criterion (Section 2.7.2)	$SIC(n) > SIC(k)$ for some k	N/A	$k_{SIC} = 2, \dots, n-2$ $SIC(k_{SIC}) = \min_{2 \leq k \leq n-2} SIC(k)$
Bayes (Section 3.6.2)	BF < 1	$p = \frac{1}{n}$	$k_{mean} = \frac{nm^0 + M_E}{m^0 + M}$ $\sum_{j=1}^{k_{med}} m_j^A = \frac{m^0 + M}{2}$ $m_{k_{mode}}^A = \max_{1 \leq k \leq n} m_k^A$
		$p = .5$	$k_{mean} = \frac{n(n-1)m^0 + M_E}{(n-1)m^0 + M}$ $\sum_{j=1}^{k_{med}} m_j^A = (n-1) \frac{m^0 + M_n}{2}$ $m_{k_{mode}}^A = \max_{1 \leq k \leq n} m_k^A$
Alternative Bayes Tests (Section 3.6.3)	$\frac{m^0}{m_k^A} < 1$ for some $k \neq n$.	$p = \frac{1}{n}$	$\hat{k} = k_{mode}$
	$\frac{m^0}{m_k^A} < \frac{1}{n-1}$ for some $k \neq n$.	$p = .5$	$\hat{k} = k_{mode}$
	$\omega(\underline{X}) = \frac{m^0}{\max_{1 \leq k \leq n-1} m_k^A} < 1$	N/A	$\hat{k} = k_{mode}$

3.7 Confidence Intervals In the Presence of a Change Point

The GenGamma-Jeffries method calculates β -level confidence intervals based on an iid random sample, $\underline{X} = \{x_i\}_{i=1}^n$. This method allows the following calculations, as discussed in previous sections:

- A confidence interval for $F(z)$: Given $z \in (0, \infty)$ calculate $U \in (0, 1)$ such that:

$$P(F(z) \geq U) = \beta. \quad (3.7.1)$$

- A confidence interval for u-quantile: $\rho_u = F^{-1}(u)$: Given $u \in (0, 1)$ calculate $Z \in (0, \infty)$ such that:

$$P(\rho_u \leq Z) = \beta \text{ where } \rho_u = F^{-1}(u) \quad (3.7.2)$$

When a random sample has a change point, k , both the values of U and Z depend on the value of k , the location of the change point. Since the calculations in (3.7.1) and (3.7.2) ignore this dependency, the calculation performed when a change point exists will introduce some error in the computed value of U or Z . The question then arises whether this error is large enough to warrant detection of change point location and further correction of the confidence intervals to take into account existence of a change point. The fact that the sample size is very small makes this question even more relevant.

In order to address these issues, we must first determine how to construct confidence intervals when a change point exists. These confidence intervals can then be compared to those derived from formulae (3.7.1) and (3.7.2).

For this purpose, we define two random variables:

$X \sim f_1(y | a_1, b_1, \lambda_1)$ and $Y \sim f_2(y | a_2, b_2, \lambda_2)$, where

- $f_1(y | a_1, b_1, \lambda_1)$ is the distribution of $\underline{X} = \{x_i\}_{i=1}^n$ through the k^{th} sample, with corresponding cdf, F_1 , and
- $f_2(y | a_2, b_2, \lambda_2)$ is the distribution of $\underline{X} = \{x_i\}_{i=1}^n$ after the k^{th} sample, with corresponding cdf, F_2 .

Then the random variable, $Z(j)$ represented by $\underline{X} = \{x_i\}_{i=1}^n$ is:

$$Z(j) = \begin{cases} X & j \leq k \\ Y & j > k \end{cases} = I(j \leq k)X + I(j > k)Y, \quad k=1,2,\dots,n$$

Given a value for k , the pdf of $Z(j)$ is of the form

$$f(z | j, k) = I(j \leq k)f_1(x) + I(j > k)f_2(y)$$

where $I(A)$ represents the indicator function of the set A and $k = n$ represents a sample without a change point. Note that here j is not a random variable. However, if we induce a uniform prior distribution on j , we will be able to use a Bayesian approach.

Hence, we assign equal probabilities to all values of j :

$$h(j) = \frac{1}{n}, \quad j = 1, \dots, n.$$

Then the joint pdfs of Z and j , and a pdf of Z are given, respectively, by the following relationships

$$f(z, j | k) = f(z | j, k)h(j) = \frac{1}{n} \left(I(j \leq k)f_1(x) + I(j > k)f_2(y) \right),$$

$$f(z | k) = \frac{1}{n} \sum_{i=1}^n \left(I(i \leq k)f_1(x) + I(i > k)f_2(y) \right)$$

Rewriting the second formula, we derive the pdf of Z given k

$$f(z) = \frac{k}{n} f_1(x) + \frac{n-k}{n} f_2(y).$$

Then, the true cumulative distribution represented by $\underline{X} = \{x_i\}_{i=1}^n$ with a change point, k is

$$F_k(z) = \frac{k}{n} F_1(z) + \frac{n-k}{n} F_2(z) \quad (3.7.3)$$

so that

$$z = F_k^{-1} \left(\frac{k}{n} F_1(z) + \frac{n-k}{n} F_2(z) \right) \quad (3.7.4)$$

Formula (3.7.3) provides an idea for calculation of the lower confidence bound for the cdf $F_k(z)$. Namely, let U_1 and U_2 be lower confidence bounds for $F_1(z)$ and $F_2(z)$, respectively. Then W_k of the following form will provide the lower confidence bound for $F_k(z)$:

$$W_k = \frac{k}{n}U_1 + \frac{n-k}{n}U_2. \quad (3.7.5)$$

Recall that here U_1 and U_2 are independent and note that W_k is different from the lower bound W which is constructed without taking into account change point k , i.e. $W_k \neq W$. Then the error in (3.7.1) is caused by using $U = W$ instead of using W_k satisfying (3.7.5).

Similarly, for a given value u and change point k , let Z be an upper bound for the quantile $\rho_U = F^{-1}(u)$ calculated without taking into account change point k and Z_{W_k} be an upper bound for $\rho_{W_k} = F_k^{-1}(W_k = u)$. Then, since $F(z) \neq F_k(z)$, we have $Z \neq Z_{W_k}$. Thus, the error in (3.7.2) is caused by using the random variable Z instead of Z_{W_k} given by (3.7.4).

3.7.1 Confidence Intervals for F(z)

Formula (3.7.5) provides basic guidelines on how to construct a lower confidence bound for distribution function $F_k(z)$ with the confidence level β . Choose any $a \in [0,1]$ and construct lower confidence bounds U_1 and U_2 for $F_1(z)$ and $F_2(z)$ based on observations $\underline{X}_1^k = \{x_i\}_{i=1}^k$ and $\underline{X}_{k+1}^n = \{x_i\}_{i=k+1}^n$, and corresponding to confidence levels β^a and β^{1-a} , respectively. Note that for any z and any independent random variables U_1 and U_2 , the following inequality is valid.

$$P\left(F_k(z) \geq \frac{k}{n}U_1 + \frac{n-k}{n}U_2\right) \geq P(F_1(z) \geq U_1)P(F_2(z) \geq U_2).$$

This inequality is true since U_1 and U_2 are independent and since the events $F_1(z) \geq U_1$

and $F_2(z) \geq U_2$ imply that $F_k(z) \geq \frac{k}{n}U_1 + \frac{n-k}{n}U_2$ by formula (3.7.5).

Since $P(F_1(z) \geq U_1) = \beta^a$ and $P(F_2(z) \geq U_2) = \beta^{1-a}$, the last formula implies that

$$P\left(F_k(z) \geq \frac{k}{n}U_1 + \frac{n-k}{n}U_2\right) \geq \beta. \text{ Hence,}$$

$$W_k = \frac{k}{n}U_1 + \frac{(n-k)}{n}U_2 \tag{3.7.6}$$

is the lower confidence bound for $F_k(z)$ with the confidence level β .

In our subsequent Monte Carlo simulations we used the value $a = 0.5$ but any $a \in [0,1]$ can be used in this calculation. For example, $a = 0$ will correspond to construction of the confidence bound on the basis of the portion of the sample after the change point only.

The confidence bounds U_1 and U_2 are constructed using the same Bayesian technique on the basis of noninformative priors as in Section 3.4.

3.7.2 Confidence Intervals for Quantile Z

In this section we shall construct Z such that $P(F_k^{-1}(u) \leq Z) \geq \beta$. To solve the problem of construction of confidence intervals for $F_k^{-1}(u)$, we let $q_1(z_1 | u)$ and $q_2(z_2 | u)$ be posterior pdfs of $F_1^{-1}(u)$ and $F_2^{-1}(u)$, respectively. The posterior pdf based on a random sample of size n is given by equation (3.3.6). However, in this case, the two pdfs are based on two independent sub-samples, the first k observations $\underline{X}_k = \{x_i\}_{i=1}^k$ and the last $n-k$ observations, $\underline{X}_{n-k} = \{x_i\}_{i=k+1}^n$. Then the joint pdf of $F_1^{-1}(u)$ and $F_2^{-1}(u)$ is:

$$q(z_1, z_2 | u) = q_1(z_1 | u)q_2(z_2 | u).$$

If $Q(z_1, z_2 | u) = Q_1(z_1 | u)Q_2(z_2 | u)$ is the joint cdf of $F_1^{-1}(u)$ and $F_2^{-1}(u)$, then the goal is to find Z such that

$$Q(Z, Z | u) = Q_1(Z | u)Q_2(Z | u) = \beta. \quad (3.7.7)$$

Then,

$$P\left(\frac{k}{n}F_1(z) + \frac{n-k}{n}F_2(z) \geq u\right) \geq P(F_1(z) \geq u)P(F_2(z) \geq u) = \beta$$

so that for a specific value of Z , this is equivalent to:

$$P(F_k(Z) \geq u) \geq \beta \Leftrightarrow P(F_k^{-1}(u) \leq Z) \geq \beta.$$

Note that we can always find the value Z satisfying equation (3.7.7) since function $Q(Z, Z | u)$ is a strictly increasing function of Z . Recall equation (3.3.8), to find confidence intervals based on a given $u \in [0, 1]$ and a random sample of size n , then

$$Q_1(Z | u) = \int_0^z \pi_p(\rho_u) d\rho_u \quad \text{where } \rho_u = F_1^{-1}(u) \quad \text{and the random sample is } \underline{X}_k = \{x_i\}_{i=1}^k$$

$$Q_2(Z | u) = \int_0^z \pi_p(\rho_u) d\rho_u \quad \text{where } \rho_u = F_2^{-1}(u) \quad \text{and the random sample is } \underline{X}_{n-k} = \{x_i\}_{i=k+1}^n$$

So we approximate Z in (3.7.7) by using the product of these functions. Then Z is the upper confidence bound for $F_k^{-1}(u)$ with confidence level β .

3.8 Error in Confidence Intervals When Change Point is Ignored

Since the calculations in (3.7.1) and (3.7.2) ignore change point existence, the calculation performed when a change point exists will introduce some error in the computed value of u or z . The question arises as to the value of such an error, and the value's dependency on the location of the change point. Further, is the error large enough to warrant detection and calculation of a change point when the sample size is small?

3.8.1 Confidence Intervals for the CDF

Let U be the confidence bound computed by ignoring the change point in equation (3.7.1), and W_k be the lower confidence bound for $F_k(z)$ computed from equation (3.7.6). Define the error in the lower bound for $F_k(z)$ as:

$$\varepsilon_U(k) = U - W_k \quad (3.8.1)$$

where the sign of ε_U determines whether U overestimates or underestimates W_k . For fixed n , the error depends on the value of k .

Since W_k is the lower confidence bound for $F_k(z)$ with the confidence level β , let α be the confidence level achieved by using the value U as the lower confidence bound for $F_k(z)$. So we have $P(F_k(z) \geq W_k) = \beta$ and $P(F_k(z) \geq U) = \alpha$. Then if the error is positive, we have $U > W_k$, thus $\alpha < \beta$, and our confidence level has decreased by ignoring the change point. If the error is negative, then $U < W_k$, thus $\alpha > \beta$ and our confidence level has increased by ignoring the change point.. Clearly, for such negative error, the value of $F_k(z)$ will be included in both confidence intervals. So that for negative error, we have improved our confidence level by ignoring the change point. However, for positive error, it is not clear that the value of $F_k(z)$ is necessarily included in the reduced confidence level interval.

The simulation models used in Section 4 provide positive errors most of the time, however, the confidence intervals with U as a lower bound always included $F_k(z)$. Thus, while we achieve a reduced confidence level, the confidence interval remained valid.

3.8.2 Confidence Intervals for the Quantile

Let Z be the value computed by ignoring the change point in equation (3.7.2), and Z_k be the upper confidence bound for $F_k^{-1}(u)$ computed from equation (3.7.7). Define the error in cdf confidence interval lower bound calculations as:

$$\varepsilon_Z(k) = Z - Z_k \quad (3.8.2)$$

where the sign of ε_Z determines whether the Z estimate was too large or too small. For fixed n , the error depends on the value of k .

Since Z_k is the upper confidence bound for $F_k^{-1}(u)$ with the confidence level β , let α be the confidence level achieved by using the value Z as the upper confidence bound for $F_k^{-1}(u)$. So we have $P(F_k^{-1}(u) \leq Z_k) = \beta$ and $P(F_k^{-1}(u) \leq Z) = \alpha$. Then if the error is positive, we have $Z > Z_k$, thus $\alpha > \beta$, and our confidence level has increased by ignoring the change point. If the error is negative, then $Z < Z_k$, thus $\alpha < \beta$ and our confidence level has decreased by ignoring the change point.. Clearly, for positive error,

the actual value of $F_k^{-1}(u)$ will be included in both confidence intervals. So that for positive error, we have improved our confidence level by ignoring the change point. However, for negative error, it is not clear that the actual value of $F_k^{-1}(u)$ is included in the reduced confidence level interval.

Simulations to investigate the error and coverage behavior of Z and Z_k will be performed at a later date for inclusion in a publication after the dissertation is completed.

4.0 SIMULATIONS

In this chapter, we shall compare the confidence intervals constructed in Chapter 3 with the nonparametric confidence intervals described in Section 2.5. We shall carry out the small sample comparison of the intervals in terms of their average length and coverage probability on the basis of Monte Carlo simulations. Also in this chapter we report simulations that investigate the effectiveness of the change point detection and location techniques described in Section 3.6. After this we shall perform simulations with selected models to investigate construction of confidence intervals when a sample contains a change point. The impact of ignoring the change point is also investigated through these simulations. All simulations are performed using a combination of Mathematica[®] and Matlab[®] software.

In order to use Monte Carlo simulations for the purposes listed above, we generate random samples from some selected Generalized Gamma distributions, as shown in (2.1.4). Section 4.1 describes this process. For each of the samples, we construct interval estimators based on equation (3.3.4) as well as the Jeffreys interval and the Second Order Corrected Interval described in Section 2.5. Section 4.2 describes these computations and provides the results of the simulations in terms of the interval lengths and coverage probabilities. Without loss of generality, in construction of confidence intervals we use $\beta=.95$.

For the change point analysis simulations, Section 4.3 describes the process and results of testing the effectiveness of the techniques for change point detection and location as described in Section 3.6. Section 4.4 describes the process and results of generating confidence intervals by taking into account or not taking into account change point location information.

4.1 Random Sample Generation

For the Monte Carlo simulations of confidence interval generation, assuming an i.i.d. sample without a change point, we generate $m=1000$ samples from the Generalized Gamma distribution for each of three choices of parameters resulting in three different Generalized Gamma distributions.

1) Distribution 1: $a=5$ $b=3$ and $\lambda=1$

$$f(x_i | 5, 3, 1) = \frac{1}{\Gamma(3)5^3} x_i^{3-1} e^{-\left(\frac{x_i}{5}\right)} = \frac{1}{250} x_i^2 e^{-\left(\frac{x_i}{5}\right)} . \quad (4.1.1)$$

2) Distribution 2: $a=1$ $b=2$ and $\lambda=.25$,

$$f(x_i | 1, 2, .25) = \frac{1}{\Gamma(2)1^2} x_i^{\frac{1}{4}(2)-1} e^{-\left(\frac{x_i^{.25}}{1}\right)} = x_i^{-.5} e^{-x_i^{.25}} . \quad (4.1.2)$$

3) Distribution 3: $a=30$ $b=.5$ and $\lambda=2$,

$$f(x_i | 30, .5, 2) = \frac{1}{\Gamma(.5)\sqrt{30}} x_i^{2(.5)-1} e^{-\left(x_i^2/30\right)} = \frac{1}{\sqrt{30\pi}} e^{-\left(x_i^2/30\right)}. \quad (4.1.3)$$

To generate a random sample from the Generalized Gamma distribution, we generate random samples with the Gamma(a,b) distribution and then apply the transformation $Y = X^{1/\lambda}$, resulting in the random sample for the i^{th} distribution, $i=1,2,3$. For each of these three distributions, we performed $m=1000$ simulation runs with the sample sizes $n=10$, $n=20$ or $n=40$.

4.2 Confidence Intervals Without Change Points

We now construct 95% lower confidence bounds, U , for $F(z)$ and 95% upper confidence bounds, Z , for $\rho_u = F^{-1}(u)$ using the procedures described in Section 3.4, and specifically using the posterior distributions described by equations (3.3.7) and (3.3.8). In the case of construction of lower confidence bounds for U , we compare confidence intervals derive in the present dissertation (which we shall refer to as GenGam) with nonparametric confidence intervals based on Jeffreys prior, equation (2.5.1), and the nonparametric second order corrected interval, equation (2.5.2), (which we shall name NPJef and NPSO, respectively). In the case of the upper confidence bound for

$\rho_u = F^{-1}(u)$, we have not compared our intervals with any benchmark intervals since we are not aware of nonparametric intervals constructed for low sample sizes.

All confidence intervals for $F(z)$ were constructed for the given z 's that correspond to the 90%, the 95% and the 99% quantiles of each of the three Generalized Gamma distributions. All confidence intervals for $F^{-1}(u)$ were constructed for the given u 's equal to 90%, 95% and 99%. Recall that $\beta = .95$ confidence level was used in all computations.

Results of simulations are summarized in Tables 2 and 3. "Average coverage" in both tables is calculated as the percentage of the intervals covering the actual value of the parameter divided by 1000, the number of simulation runs. In both Tables 2 and 3, the goal is 95% coverage.

Table 2: Lower Confidence Bounds, U, for $u=F(z)$, 1000 Simulations

Distribution 1: $a=5$, $b=3$, and $\lambda=1$								
			Average Interval Length 1-U			Average Coverage (%)		
u	z	n	GenGam	NPJef	NPSO	GenGam	NPJef	NPSO
0.90	26.61	40	0.1819	0.2056	0.2065	94.20	91.00	91.00
		20	0.2300	0.2686	0.2705	93.10	89.50	89.50
		10	0.2832	0.3136	0.3206	97.50	100.0	100.0
0.95	31.48	40	0.1075	0.1421	0.1434	95.10	84.80	84.80
		20	0.1432	0.1745	0.1782	98.00	100.0	100.0
		10	0.2043	0.2464	0.2566	100.0	100.0	100.0
0.99	42.03	40	0.0458	0.0656	0.0683	100.0	100.0	100.0
		20	0.0685	0.1077	0.1135	100.0	100.0	100.0
		10	0.1290	0.1874	0.2009	100.0	100.0	100.0

Distribution 2: $a=1$, $b=2$, and $\lambda=0.25$								
			Average Interval Length 1-U			Average Coverage (%)		
u	z	n	GenGam	NPJef	NPSO	GenGam	NPJef	NPSO
0.90	228.9	40	0.1907	0.2079	0.2089	92.10	92.30	92.30
		20	0.2342	0.2630	0.2650	91.70	87.30	87.30
		10	0.2965	0.3140	0.3209	98.70	100.0	100.0
0.95	506.4	40	0.1116	0.1424	0.1437	92.50	85.90	85.90
		20	0.1446	0.1719	0.1758	97.90	100.0	100.0
		10	0.2150	0.2467	0.2567	100.0	100.0	100.0
0.99	1941	40	0.0420	0.0652	0.0678	100.0	100.0	100.0
		20	0.0694	0.1087	0.1145	100.0	100.0	100.0
		10	0.1379	0.1872	0.2008	100.0	100.0	100.0

Distribution 3: $a=30$, $b=0.5$, and $\lambda=2$								
			Average Interval Length 1-U			Average Coverage (%)		
u	z	n	GenGam	NPJef	NPSO	GenGam	NPJef	NPSO
0.90	6.37	40	0.1838	0.2039	0.2048	95.50	92.10	92.10
		20	0.2277	0.2687	0.2706	97.20	88.50	88.50
		10	0.2924	0.2965	0.3263	99.70	100.0	100.0
0.95	7.59	40	0.1174	0.1396	0.1409	97.40	86.30	86.30
		20	0.1590	0.1766	0.1803	99.70	100.0	100.0
		10	0.2296	0.2147	0.2629	100.0	100.0	100.0
0.99	9.98	40	0.1155	0.1269	0.1284	100.0	100.0	100.0
		20	0.1587	0.1766	0.1803	100.0	100.0	100.0
		10	0.2296	0.2147	0.2629	100.0	100.0	100.0

Table 3: Upper Confidence Bounds, $\hat{\rho}_u$, for $\rho_u = F^{-1}(u)$, 1000 Simulations

Distribution 1: a=5, b=3, and $\lambda=1$					
u	ρ_u	n	Av.length $\hat{\rho}_u$	$F(\hat{\rho}_u)-u$	Av. Coverage (%)
0.90	26.61	40	26.67	0.06	98.10
		20	26.68	0.07	97.20
		10	26.70	0.09	96.60
0.95	31.48	40	31.51	0.03	95.65
		20	31.52	0.04	94.90
		10	31.53	0.05	95.72
0.99	42.03	40	42.04	0.01	95.05
		20	42.04	0.01	93.35
		10	42.04	0.01	91.21

Distribution 2: a=1, b=2, and $\lambda=0.25$					
u	ρ_u	n	Av.length $\hat{\rho}_u$	$F(\hat{\rho}_u)-u$	Av. Coverage (%)
0.90	228.9	40	228.99	0.07	93.70
		20	229.00	0.08	89.75
		10	229.01	0.10	88.30
0.95	506.4	40	506.48	0.04	90.80
		20	506.48	0.04	88.00
		10	506.49	0.05	85.50
0.99	1941	40	1941.97	0.01	87.80
		20	1941.97	0.01	84.00
		10	1941.97	0.01	83.50

Distribution 3: a=30, b=0.5, and $\lambda=2$					
u	ρ_u	n	Av.length $\hat{\rho}_u$	$F(\hat{\rho}_u)-u$	Av. Coverage (%)
0.90	6.37	40	6.43	0.06	96.40
		20	6.45	0.08	95.60
		10	6.47	0.10	94.80
0.95	7.59	40	7.63	0.04	97.30
		20	7.64	0.05	95.90
		10	7.64	0.05	96.00
0.99	9.98	40	9.99	0.01	100.0
		20	9.99	0.01	99.90
		10	9.99	0.01	99.20

If U is a lower confidence bound for $u=F(z)$ and $\hat{\rho}_u$ is an upper confidence bound for ρ_u constructed using observations of one simulation run, then the respective length of the confidence intervals are $1-U$ and $\hat{\rho}_u$. In Tables 2 and 3, the average lengths of these confidence intervals are calculated using only the intervals that cover the actual value of the parameter, hence, it is possible for a technique to provide simultaneously the better coverage and shorter average length of the interval. Observe also, that the length of the confidence interval for ρ_u has an upper bound that is equal to the value of ρ_u itself. Hence, if $\rho_u = 506.44$, then the average length of the confidence interval cannot be less than 506.44. For this reason, in Table 3 we put $F(\hat{\rho}_u) - F(\rho_u) = F(\hat{\rho}_u) - u$ as a measure of the quality of confidence intervals for ρ_u .

It is easy to see that the method based on the Generalized Gamma distribution delivers shorter confidence intervals with better coverage than nonparametric techniques. The reason is that it is practically impossible for a nonparametric technique to adjust to the tails of distribution F and in reliability only the tail behavior of F is of interest. If $n=10$, the confidence intervals for u are too conservative no matter what technique is chosen, however, GenGam intervals are shorter on the average.

4.3 Robustness of Technique

Since reliability data may come from various distributions, it is interesting to study how well the confidence intervals described above perform when data comes from a distribution other than Generalized Gamma. Yet, since the generalized gamma distribution is extremely flexible, the only case when a positive continuous random variable cannot be represented by this distribution is the case when the distribution of data is not unimodal. This is the change point case and is discussed further in Section 4.5.

It is very common to have reliability data in the form of failure counts. Very often this sort of data is represented by a Poisson processes (see e.g. Rigdon and Basu [34]) and is examined for the presence of trend. Therefore, in this section we generate i.i.d. samples from a Poisson distribution and construct confidence intervals for $F(z)$ and $F^{-1}(u)$ assuming that the data came from generalized gamma distribution, thus evaluating how robust our methodology is.

This approach, however, has an obvious limitation. A Poisson random variable can take a zero value that is impossible for a variable with generalized gamma distribution. Moreover, any $X_i = 0$ turns P_n into zero, making further analysis impossible. Nevertheless, if data comes from a Poisson(θ) distribution with parameter θ being fairly

large and the sample size is relatively small, one is unlikely to see zero values in a sample.

In what follows, we generate data from three Poisson distributions with $\theta = 5, 10$ and 25 , respectively. We construct confidence intervals for $F(z)$ and $F^{-1}(u)$ and compare the confidence intervals for U with the nonparametric intervals exactly in the same manner as in Section 4.2. Comparisons are carried out on the basis of $m = 500$ simulation runs. Results of simulations are presented in Tables 4 and 5.

Table 4: Lower Confidence Bounds,U, for $u=F(z)$, Poisson Data, 1000 Simulations

Poisson(θ) data with $\theta = 5$								
			Average Interval Length 1-U			Average Coverage (%)		
u	z	n	GenGam	NPJef	NPSO	GenGam	NPJef	NPSO
0.90	8.00	40	0.1721	0.1818	0.1731	88.10	71.40	71.40
		20	0.2151	0.2230	0.2253	95.20	75.40	75.40
		10	0.2947	0.3017	0.3113	90.80	88.20	88.20
0.95	9.00	40	0.1104	0.1212	0.1226	89.80	65.40	65.40
		20	0.1396	0.1444	0.1489	95.20	100.0	100.0
		10	0.1588	0.1744	0.1801	92.80	85.00	85.00
0.99	11.00	40	0.0435	0.0571	0.0599	100.0	100.0	100.0
		20	0.0708	0.1010	0.1071	100.0	100.0	100.0
		10	0.1037	0.1152	0.1224	100.0	100.0	100.0

Poisson(θ) data with $\theta = 10$								
			Average Interval Length 1-U			Average Coverage (%)		
u	z	n	GenGam	NPJef	NPSO	GenGam	NPJef	NPSO
0.90	14.00	40	0.1843	0.1873	0.1853	91.40	86.20	86.20
		20	0.2360	0.2461	0.2182	92.00	82.40	82.40
		10	0.2743	0.2979	0.3053	90.60	89.20	89.20
0.95	15.00	40	0.1253	0.1289	0.1303	94.80	81.80	81.80
		20	0.1692	0.1672	0.1713	96.40	100.0	100.0
		10	0.2004	0.2433	0.2688	95.60	87.40	87.40
0.99	18.00	40	0.0406	0.0599	0.0626	100.0	100.0	100.0
		20	0.0719	0.1044	0.1104	100.0	100.0	100.0
		10	0.1132	0.1367	0.1477	100.0	100.0	100.0

Poisson(θ) data with $\theta = 25$								
			Average Interval Length 1-U			Average Coverage (%)		
u	z	n	GenGam	NPJef	NPSO	GenGam	NPJef	NPSO
0.90	32.00	40	0.1635	0.1681	0.1692	92.40	78.80	78.80
		20	0.2059	0.2367	0.2388	93.00	73.00	73.00
		10	0.2246	0.2542	0.2684	91.60	81.40	81.40
0.95	33.00	40	0.1218	0.1357	0.1370	98.40	84.20	84.20
		20	0.1594	0.1606	0.1648	100.0	100.0	100.0
		10	0.1794	0.1975	0.2383	97.80	82.80	82.80
0.99	37.00	40	0.0410	0.0663	0.0689	100.0	100.0	100.0
		20	0.0704	0.1026	0.1087	100.0	100.0	100.0
		10	0.1253	0.1586	0.1877	100.0	100.0	100.0

Table 5: Upper Confidence Bounds, $\hat{\rho}_u$, for $\rho_u = F^{-1}(u)$, Poisson Data, 1000 Simulations

Poisson(θ) data with $\theta = 5$					
u	ρ_u	n	Av.length $\hat{\rho}_u$	$F(\hat{\rho}_u)$ -u	Av. Coverage (%)
0.90	8.00	40	9.082	0.0400	95.80
		20	9.847	0.0690	94.40
		10	10.431	0.0690	92.80
0.95	9.00	40	10.322	0.0190	94.80
		20	11.285	0.0400	93.20
		10	12.815	0.0480	91.60
0.99	11.00	40	12.987	0.0080	93.80
		20	14.354	0.0094	91.80
		10	15.836	0.0094	91.80

Poisson(θ) data with $\theta = 10$					
u	ρ_u	n	Av.length $\hat{\rho}_u$	$F(\hat{\rho}_u)$ -u	Av. Coverage (%)
0.90	14.00	40	15.788	0.0520	97.02
		20	16.731	0.0730	96.00
		10	17.810	0.0900	95.20
0.95	15.00	40	17.314	0.0230	95.80
		20	18.358	0.0430	96.20
		10	19.250	0.0430	93.60
0.99	18.00	40	20.677	0.0090	94.00
		20	22.323	0.0094	93.80
		10	24.330	0.0099	93.60

Poisson(θ) data with $\theta = 25$					
u	ρ_u	n	Av.length $\hat{\rho}_u$	$F(\hat{\rho}_u)$ -u	Av. Coverage (%)
0.90	32.00	40	34.283	0.0510	98.40
		20	35.537	0.0780	97.4
		10	36.230	0.0780	95.6
0.95	33.00	40	36.608	0.0400	100.0
		20	37.862	0.0410	98.02
		10	38.570	0.0450	96.40
0.99	37.00	40	40.673	0.0080	95.60
		20	42.626	0.0094	95.60
		10	44.110	0.0097	93.80

Again, the method based on the Generalized Gamma distribution delivers shorter confidence intervals with better coverage than nonparametric techniques in spite of the fact that the data *not* come from this distribution. The reason perhaps lies in the relationship between the Poisson and Gamma distributions. If $F_{Poi}(x|\theta)$ and $F_G(x|1,b)$ are cdfs of the Poisson(θ) distribution and gamma distribution with the unit scale parameter and shape parameter b , then $F_{Poi}(b-1|\theta) = 1 - F_G(\theta|1,b)$ (see e.g. Casella and Berger [1], page 130).

Since confidence intervals based on the Generalized Gamma distribution do not require specification of parameters, confidence intervals for Poisson data are well approximated by generalized gamma distribution and are free from the "curse of discreteness" which is reported in relation to binomial data (see e.g. Agresti and Coull [40] or Brown, Cai and DasGupta [41]).

4.4 Change Point Detection and Location Techniques

Section 3.6 describes five techniques for change point detection and location, summarized in Table 1. In addition, a benchmark technique based on the Schwartz Information Criterion, is described in Section 2.7.2. This benchmark technique is recommended by Chen and Gupta [35] for change point detection and location. This

section provides simulations for comparison of effectiveness for each of these six techniques based on the small sample size of $n = 60$. This sample size allows a reasonable range of subsample sizes as small as 10, as k varies between 10 and 50. Prior simulation sections have shown that the GenGamma-Jeffreys method is effective for confidence interval estimation with a sample size as small as $n = 10$.

There are three cases of interest for change point detection and location:

- First is when the change in distribution causes a minor difference between the mean and variance of the two distributions. In such a case, the change point may be more difficult to detect and locate.
- Second is when the change in distribution causes a significant difference between the mean and variance of the two distributions. In such a case, the change point should be easier to detect and locate.
- Third case is when no change point exists. Then the testing technique should reject existence of a change point.

To examine each of the above three cases, three models were selected. Model #1 represents a change point in a random sample changing from the Generalized Gamma distribution with $a=5$, $b=3$, and $\lambda=5$ to the Generalized Gamma distribution with $a=7$, $b=5$, and $\lambda=6$. This slight distribution difference can be seen from the following graph, where the initial distribution is the solid line.

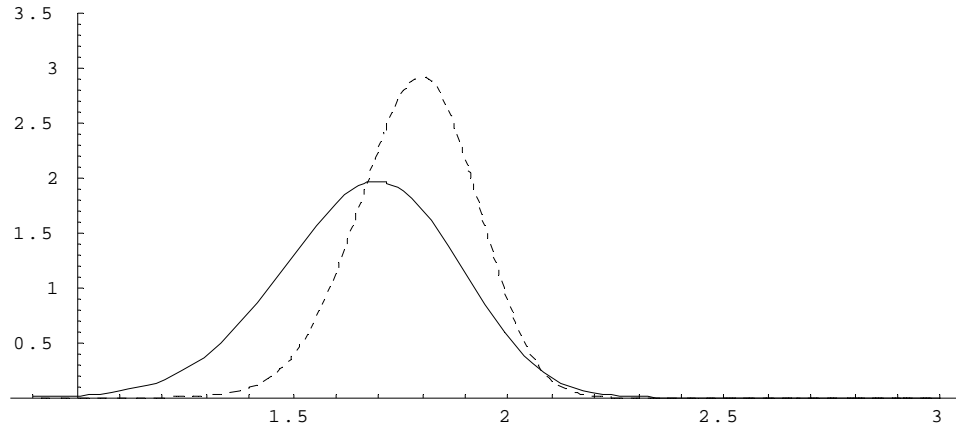


Figure 4: Model 1

Model #2 represents a change point in a random sample changing from the Generalized Gamma distribution with $a=7$, $b=5$, and $\lambda=6$ to the Generalized Gamma distribution with $a=5$, $b=20$, and $\lambda=6$. This distribution difference can be seen from the following graph, where the initial distribution is the solid line. The mean and variance difference for this model is larger than in Model #1.

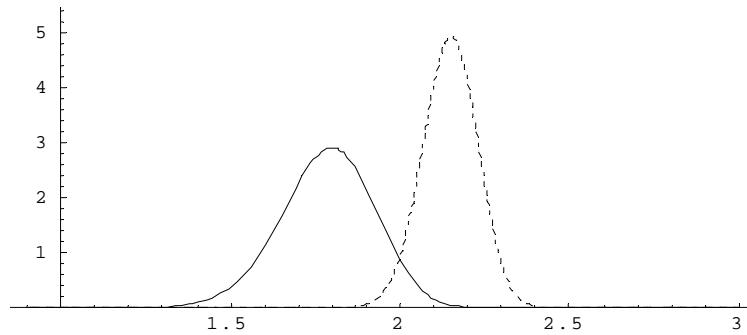


Figure 5: Model 2

Model #3, used to test detection accuracy when no change point exists, is a set of randomly generated Generalized Gamma random samples.

For all three models, accuracy of change point detection is examined. For the first two models the accuracy of estimation of location k is evaluated as well. Calculations are performed for each of the techniques and corresponding estimators listed in Table 1. For each of the three models, 1000 random samples of size $n=60$ were generated. For the change point detection simulations using Models 1 and 2, a change point is inserted randomly at a location k in $[10,50]$, to allow the first and second sub-samples to have reasonable size for later confidence interval estimation. Location k has a discrete uniform distribution in $[10,50]$. For Model 3, no change point is inserted. Detection techniques are then tested to see if a change point is detected when one does exist (Models 1 and 2) or if a change point is detected when none exists (Model 3). For the change point location estimator accuracy, each simulation is only for Models 1 and 2 and includes each value of k in $[10,50]$. For each k value, the distance between the estimator and actual k value is measured, whether the change point was actually detected, or not.

Table 6: Change Point Detection Accuracy (%)

Method	Hypothesis Test (Reject H_0)	P Estimator	Change Point Detected Model 1	Change Point Detected Model 2	No Change Point Detected Model 3
Schwartz Information Criterion (Section 2.7.2)	SIC(n) > SIC(k) for some k	N/A	59.5	67.2	15.1
Bayes (Section 3.6.2)	BF < 1	$p = \frac{1}{n}$	100	100	0
		$p = .5$	100	100	0
Alternative Bayes Tests (Section 3.6.3)	$\frac{m^0}{m_k^A} < 1$ for some $k \neq n$.	$p = \frac{1}{n}$	100	100	0
	$\frac{m^0}{m_k^A} < \frac{1}{n-1}$ for some $k \neq n$.	$p = .5$	100	100	0
	$\omega(\underline{X}) = \frac{m^0}{\max_{1 \leq k \leq n-1} m_k^A} < 1$	N/A	85.3	99.9	48.5

Results of change point detection simulations are listed in Table 6. The accuracy value listed in the table represents the percentage of correct detections per 1000 simulations. As can be seen from Table 6, an accurate assessment of the absence of the change point represents greater challenge than detection of the change point when it indeed exists. Accuracy detection for Model 3, where no change point exists, is far lower than that for Models 1 and 2, where a change point exists. Bayes methods (lines 2 and 3 of Table 6)

and Alternative Bayes Tests (lines 4 and 5) are practically useless in this regards since they detect the change point every time, whether it is present or not. The Benchmark method, SIC, is only accurate 15% of the time. However the k_{mode} test (on line 6 of Table 4) has 48.5% accuracy in detection that no change point exists. This test also provides better accuracy in detection of change point than SIC in the cases when a change point exists. Hence, this test is preferable to the other tests described in Section 3.6.

The next property to be examined for each technique is the accuracy of the estimation of the change point location. We conduct comparison for estimators listed in Table 1. Results of these simulations are presented in Table 7. Accuracy of the estimator is measured as the mean distance from the actual change point over 1000 simulations. For Bayes estimators, the accuracies of the estimators based on the mean, the median and the mode are reported. For first two types of estimators based on Alternative Bayes Tests we provide the mode only, because the test was based on this estimator. For the SIC and k_0 tests, we provide two kinds of results. The row marked “correct” indicates the accuracy of a change point location when the change point is correctly determined. The row marked “all” reports the accuracy of the estimators of k for all simulation runs, even those where the change point was not correctly detected. If the change point is correctly detected in 100% of cases, then ‘all’=‘correct’.

Table 7: Change Point Estimates

Method	Hypothesis Test (Reject H_0)	P Estimator	K Estimator		Accuracy Model 1	Accuracy Model 2
Schwartz Information Criterion (Section 2.7.2)	SIC(n) > SIC(k) for some k	N/A	k_{SIC}	all	19.83	23.47
				correct	24.92	20.43
Bayes (Section 3.6.2)	BF<1	$p = \frac{1}{n}$	k _{mean}		8.38	.75
			k _{med}		8.09	.57
			k _{mode}		9.79	.27
		$p = .5$	k _{mean}		9.86	.89
			k _{med}		27.24	27.24
			k _{mode}		10.22	.47
Alternative Bayes Tests (Section 3.6.3)	$\frac{m^0}{m_k^A} < 1$ for some $k \neq n$.	$p = \frac{1}{n}$	k _{mode}		9.79	.27
	$\frac{m^0}{m_k^A} < \frac{1}{n-1}$ for some $k \neq n$.	$p = .5$	k _{mode}		10.22	.47
	$\omega(\underline{X}) = \frac{m^0}{\max_{1 \leq k \leq n-1} m_k^A} < 1$	N/A	k _{mode}	all	9.79	.27
				correct	10.53	.25

Note that the estimator k_{mode} for the test on Line 6 provides similar results to the estimator k_{mode} for $p = \frac{1}{n}$, and these estimators have the best accuracy for Model 2, where a change point is easier to detect. However, the best accuracy for Model 1 is

achieved by estimator k_{med} for $p = \frac{1}{n}$, where change point location is harder to detect.

So, once a change point is detected, any of these four techniques could be applied towards estimating a change point location.

4.5 Confidence Intervals with Change Points

After a change point is detected and located, the goal is to generate confidence intervals for $F(z)$ taking existence of a change point into account, using the techniques of Section 3.7.1, and compare these results to the confidence interval estimates when existence of a change is ignored, and construction of confidence intervals is based on equation (3.7.1). We will also compute the error given by equations (3.8.1).

Because we are studying confidence intervals for the tail of distribution $F(z)$, when the tails of both distributions $F_1(z)$ and $F_2(z)$ are almost identical, as in Model 1, the error will clearly be small. Of greater interest are models like Model 2, shown below in Figure 6, with significant differences in the tails of $F_1(z)$ and $F_2(z)$. Recalling that Model 2 is the Generalized Gamma distribution with $a=7$, $b=5$, and $\lambda=6$ as the first pdf f_1 , changing to the Generalized Gamma distribution with $a=5$, $b=20$, and $\lambda=6$, which is called f_2 . The dotted line in Figure 6 is f_2 .

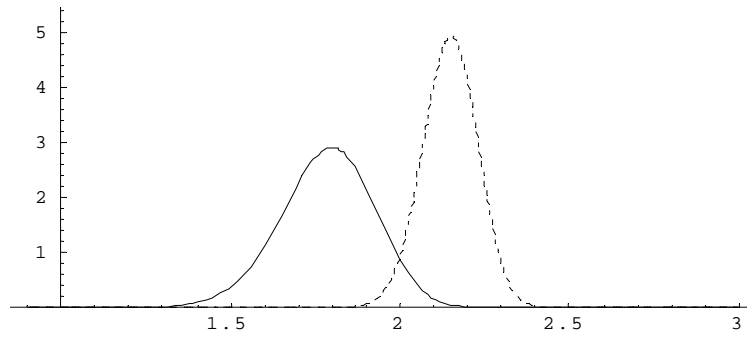


Figure 6: Model 2

In addition, Model 2R shown in Figure 7 represents the reverse of Model 2, which is a transition from f_2 to f_1 . The dotted line in Figure 7 is f_2 .

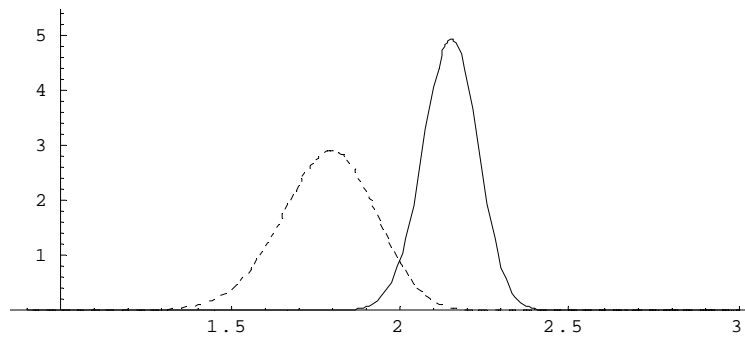


Figure 7: Model 2R

Monte Carlo simulations will use 25 samples from Model and 25 samples from Model 2R to investigate the effect of the location of the change point on the error magnitude. Simulations are conducted with a total sample size of $n = 60$ and confidence level

$\beta = .95$. For each random sample, a change point, k , between 11 and 50 is inserted. Thus there are a total of 40 simulations for each random sample, resulting in 1000 total simulations for ε_U in Model 2 and 1000 total simulations for ε_U in Model 2R,. In each of those cases, 95% confidence intervals are generated

For each simulation, confidence bound W_k is constructed using equation (3.7.6). Equation (3.7.1) will be used to calculate a confidence bound U , derived by ignoring the change point. The error given in (3.8.1) is then calculated. For both Model 2 and 2R, F_1 represents the initial distribution and F_2 is the distribution after the change point.

For the $F(z)$ confidence interval simulations, we will use $\alpha = 0.5$. Thus, we are simulating finding U_1 and U_2 such that $P(F_1(z) \geq U_1) = \sqrt{\beta}$ and $P(F_2(z) \geq U_2) = \sqrt{\beta}$.

For each k value simulation, we will find $W_k = \frac{k}{n}U_1 + \frac{(n-k)}{n}U_2$, the lower confidence

bound for $F_k(z)$ with the confidence level $\beta = .95$. For each value of k , the given z will

be $z_k = \frac{k}{n}z_1 + \frac{n-k}{n}z_2$ where $1 - F_1(z_1) = \sqrt{\beta}$ and $1 - F_2(z_2) = \sqrt{\beta}$, which are computed

directly from the known F_1 and F_2 distributions. Since this is what is being simulated,

the actual value for $F_k(z)$ is $F_k(z) = \frac{k}{n}F_1(z_k) + \frac{n-k}{n}F_2(z_k)$ for each k value simulation.

The results of $F(z)$ confidence interval simulations for Model 2 are shown in Figure 8, a plot of the change point value k versus mean ε_U at k , and Figure 9 which demonstrates

coverage of $F_k(z)$ by showing the mean of computed values of U and W_k as compared to $F_k(z)$ for each k . Similarly, Figures 10 and 11 are the equivalent charts for Model 2R.

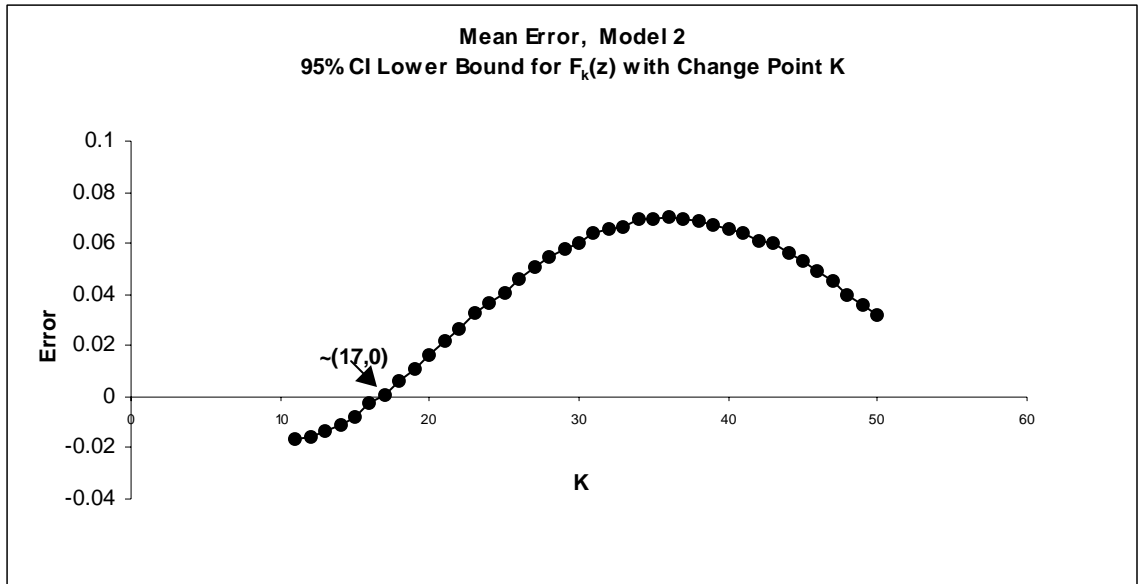


Figure 8: ε_U vs. k

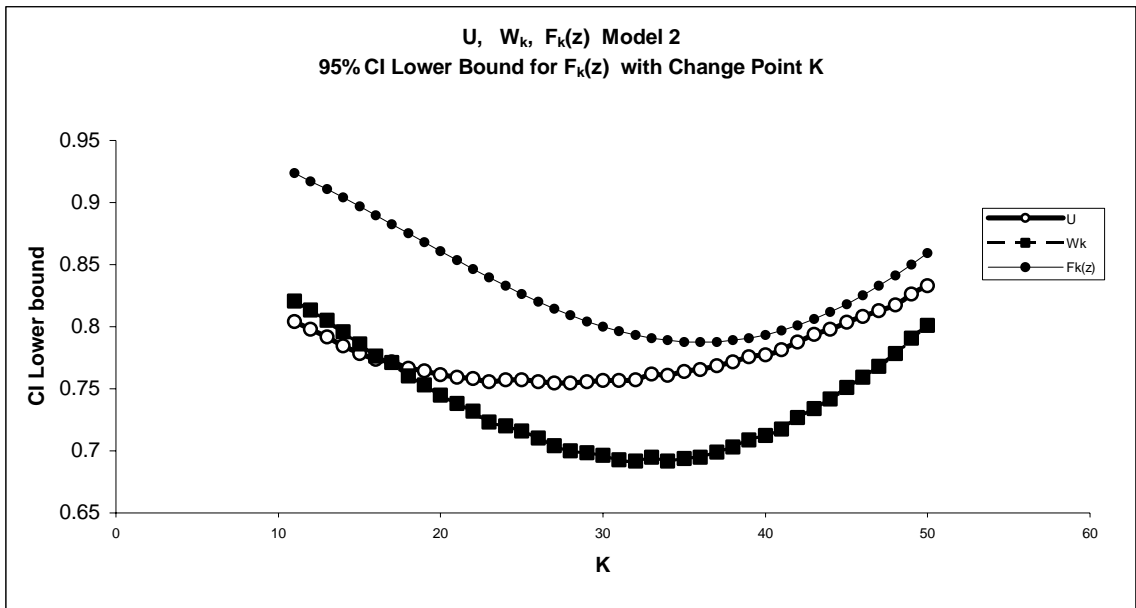


Figure 9: U , W_k , $F_k(z)$ vs. k

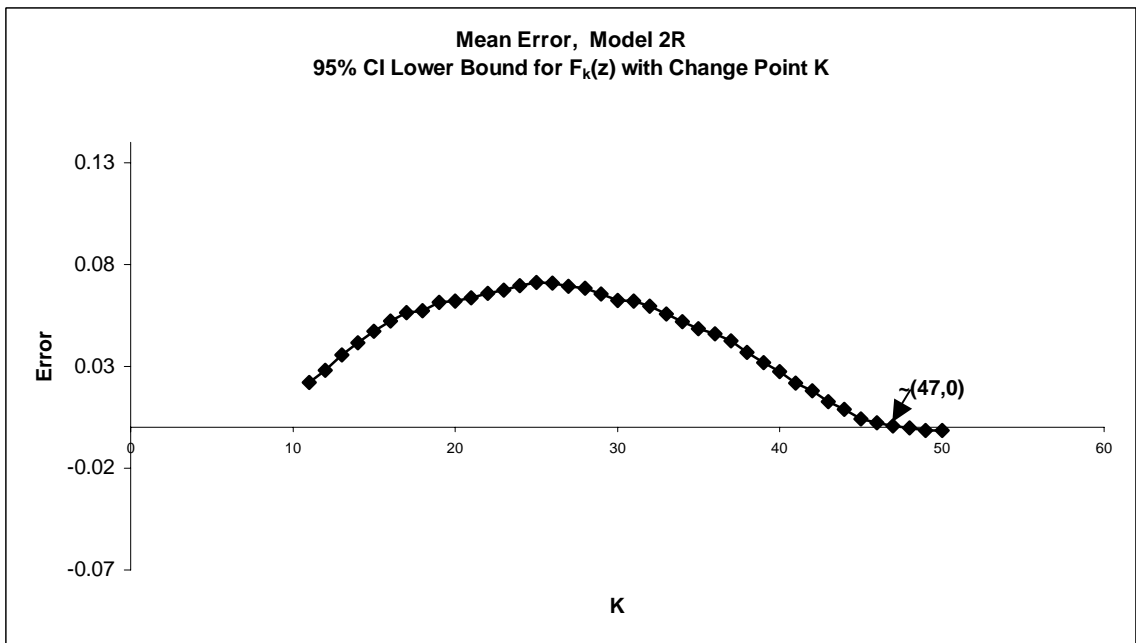


Figure 10: ε_U vs. k

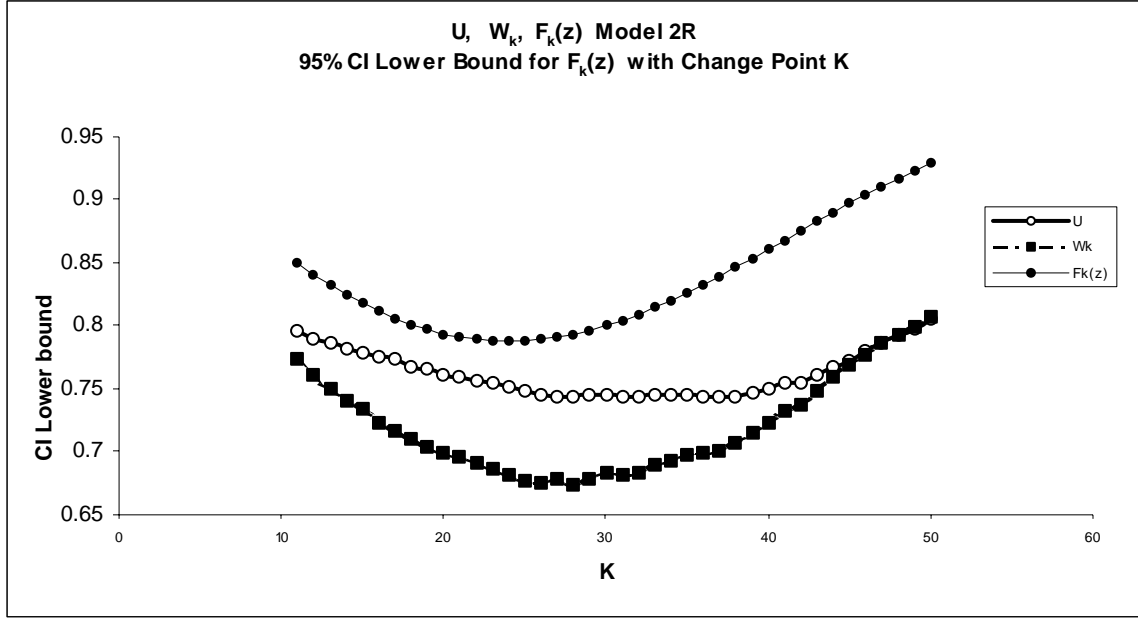


Figure 11: $U, W_k, F_k(z)$ vs. k

Clearly from these charts, the error generated is usually positive for both models, indicating there is a possibility that $F_k(z)$ will not be covered by the confidence interval that ignores existence of the change point. However, for these models the error was small enough that in both cases, $F_k(z)$ was indeed included by the defined confidence interval that ignored the change point (see Figures 9 and 11). Thus, the confidence interval level of .95 was reduced slightly by the error value, nevertheless retaining coverage of the target value while ignoring the change point. It is therefore clear, that for these models and a small sample size, it is not critical to capture the existence of a change point in the distribution, and assuming a complete i.i.d. sample is probably adequate.

5.0 CONCLUSIONS

In the present dissertation we considered construction of confidence intervals for a cumulative distribution function $F(z)$ and its inverse, quantile function $F^{-1}(u)$, at some fixed points z and u on the basis of an iid sample $\underline{X} = \{x_i\}_{i=1}^n$ where n is relatively small. While construction of nonparametric confidence intervals for $F(z)$ is related to interval estimation for binomial proportion and has consequently attracted much interest, the confidence intervals for quantiles $F^{-1}(u)$ are much less explored.

In addition, confidence intervals for binomial proportion p suffer from the “curse of discreteness” which is reported in relation to binomial data (see e.g. Agresti and Coull [40] or Brown, Cai and DasGupta [41]) exhibiting inadequate coverage when p is close to zero or one. Therefore, when X is a continuous random variable, it may be a good alternative to nonparametric confidence intervals to model the sample as having a flexible Generalized Gamma distribution with all three parameters being unknown. This distribution is able to emulate a wide variety of curves, so that a majority of the distributions used in reliability or survival analysis are its particular cases. The confidence intervals were constructed on the basis of the Jeffreys noninformative prior.

To demonstrate the advantages of the method proposed in the dissertation, we first showed (by simulations) that it indeed brings significant improvement over nonparametric techniques when the data has a Generalized Gamma distribution.

Furthermore, we studied robustness of our method by applying it to data that has a distribution different from Generalized Gamma. Since the only case when a positive continuous random variable cannot be represented by this distribution is the case when the distribution of data is clearly not unimodal and since it is virtually impossible to make sure that the data is not unimodal on the basis of small number of observations, we generate a sample from a discrete distribution, namely Poisson, and construct confidence intervals for $F(z)$ and $F^{-1}(u)$ assuming that the data came from a Generalized Gamma distribution, thus evaluating how robust our methodology is. Numerical studies show that confidence intervals constructed under the (wrong) assumption that the data came from a Generalized Gamma distribution still outperform nonparametric confidence intervals. The paper based on the above material has been submitted [36] for publication.

The methods described in the first part of the dissertation are valid only if the observations are independent and identically distributed. While the first assumption is usually validated by physical independence, the second assumption (of identical distribution) is sometimes violated and this may have a critical effect on reliability. The second part of the dissertation deals with the situation when the distribution of the observations changes over time.

The change in the distribution of the sample may be gradual, and then this change is referred to as trend. However, very often change in distribution occurs instantaneously

due to some circumstances or events. The problem of detection and location of the point when this instantaneous change occurred is called the change point problem. Investigating the change point problem in relation to construction of confidence intervals for $F(z)$ and $F^{-1}(u)$ constitutes the second half of the dissertation.

We assumed that the sample still has a flexible Generalized Gamma Distribution and, due to small sample size, has at most one change point. We suggested several methods for detection and location of the change point and studied their performance via Monte Carlo simulations. We compared the techniques proposed in the dissertation to the benchmark, the Schwartz Information Criterion. We developed objective Bayesian methods for construction of confidence intervals for distribution function $F(z)$ and quantile function $F^{-1}(u)$, once the change point is detected. Further, we investigated performance of the confidence intervals for $F(z)$ via numerical simulations. Numerical study of the confidence intervals for $F^{-1}(u)$ is still in process.

The techniques developed in this dissertation can be useful for reliability studies in many fields of application, especially as a preliminary study to quickly capture areas of concern requiring further in-depth analysis. We plan to pursue continued study and development of these techniques, particularly in the less explored area of confidence intervals for quantile functions.

LIST OF REFERENCES

- [1] Casella, G. and Berger, R. (2002) *Statistical Inference*, Duxbury, 2nd Edition.
- [2] Ebeling, C. (1997) *Reliability and Maintainability Engineering*, McGraw-Hill.
- [3] Berger, R. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag.
- [4] Stacy, E. W. (1962) *A Generalization of the Gamma Distribution*, Annals of Mathematical Statistics, Vol. 33, Issue 3, 1187-1192
- [5] Liseo, B. (1993) *Elimination of Nuisance Parameters with Reference Priors*, Biometrika, Vol. 80, Issue 2, 295-304
- [6] Cai, T.T. (2003) *One-Sided Confidence Intervals in Discrete Distributions*, Technical Report, Department of Statistics, University of Pennsylvania
- [7] Brown, L.D, Cai, T. and DasGupta, A. (2002) *Confidence Intervals for a Binomial Proportion and Asymptotic Expansions*, The Annals of Statistics 30, 160-201
- [8] Jeffreys, H. (1961) *Theory of Probability*, Oxford University Press, 3rd ed.
- [9] Wolfram, S. (2000) *The Mathematica Book*, Cambridge University Press
- [10] Amoroso, L. (1925) *Ricerche intorno alla curva dei redditi*, Ann. Mat. Pura Appl, 4th Edition, Vol 21, 123-159
- [11] d'Addario, R. (1932) *Intorno alla curia dei redditi di Amoroso*, Riv. Italiana Statist., Econ. Finanza, 4th Edition, Vol. 1
- [12] Edwards, J. (1954) *A Treatise on the Integral Calculus*, Chelsea Publications, New York
- [13] Haight, F.A. (1961) *Index to the Distributions of Mathematical Statistics*, National Bureau of Standards, Sec. B 65B, No. 1, 23-60
- [14] Pearson, K. (1934) *Tables of the Incomplete Gamma Function*, Cambridge University Press
- [15] Laplace, P.S. (1812) *Theorie Analytique des Probabilites*, Courcier, Paris

- [16] Jaynes, E.T. (1983) *Papers on Probability, Statistics, and Statistical Physics*, Reprint collection, R.D. Rosenkrantz 9(ed.), Reidel, Dordrecht
- [17] Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*, Capital City Press, 6th Edition
- [18] Ascher, H. and Feingold, H. (1984) *Repairable Systems Reliability*, Marcel Dekker, Inc., New York
- [19] Rigdon, S. and Basu, A. (2000) *Statistical Methods for the Reliability of Repairable Systems*, John Wiley & Sons, Inc.
- [20] Hollander, M. and Wolfe, D. (1973) *Nonparametric Statistical Methods*, John Wiley & Sons, Inc.
- [21] Tapia, R. and Thompson, J. (1978) *Nonparametric Probability Density Estimation*, Johns Hopkins University Press Ltd., London
- [22] Collani, E. and Drager, K. (2001) *Binomial Distribution Handbook for Scientists and Engineers*, Birkhauser, Boston
- [23] Balakrishnan, R. and Rao, C. (2001) *Advances in Reliability*, Handbook of Statistics 20, Elsevier Science B.V., Amsterdam, The Netherlands
- [24] Andrews, L. (1998) *Special Functions of Mathematics for Engineers*, 2nd Edition, Oxford Science Publications, Oxford University Press
- [25] Zwillinger, D. (1996) *CRC Standard Mathematical Tables and Formulae*, 30th Edition, CRC Press
- [26] Lehman, E.L. (1983) *Theory of Point Estimation*, Wiley, New York
- [27] Agresti, A. and Coull, B.A. (1998) *Approximate is better than "exact" for interval estimation of binomial proportions*, The American Statistician 52, 119-126
- [28] Blyth, C.R. and Still, H.A. (1983) *Binomial Confidence Intervals*, J. Amer. Stat. Assoc. 78, 108-116
- [29] Cressie, N (1980) *A finely tuned continuity correction*, Ann. Inst. Statist. Math 30, 435-442
- [30] Newcombe, R.G. (1998) *Two-sided Confidence Intervals for the Single Proportion; comparison of several methods*, Statistics in Medicine 17, 857-872

- [31] Santner, T.J. (1998) *A Note on Teaching Binomial Confidence Intervals*, Teaching Statistics 20, 20-23
- [32] Vollset, S.E. (1993) *Confidence Intervals for a Binomial Proportion*, Statistics in Medicine 12, 809-824
- [33] Brown, L.D, Cai, T. and DasGupta, A. (2001) *Interval Estimation for a Binomial Proportion (with discussion)*, Statistical Science 16, 101-133
- [34] Rigdon, S. and Basu, A. (2000) *Statistical Methods for the Reliability of Repairable Systems*, Wiley Series in Probability and Statistics, Chapter 4.
- [35] Chen, J. and Gupta, A.K.(2000) *Parametric Statistical Change Point Analysis*, Birkhauser, Boston.
- [36] Pensky, M. and Heard, A. (2005) *Confidence Intervals for Reliability and Quantile Functions with Application to NASA Space Flight Data*, IEEE Transactions on Reliability (pending).
- [37] Stacy, E.W. and Mihram G.A. (1965) *Parameter Estimation for a Generalized Gamma Distribution*, *Technometrics*, Vo.7, No. 3, 349-358.
- [38] Abramowitz, M., Stegun, I.A. (1992) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Reprint of the 1972 edition. Dover Publications, New York.
- [39] Gradshteyn, I.S., and Ryzhik, I.M. (1980) *Tables of Integrals, Series and Products*, Academic Press, New York.
- [40] Agresti, A., Coull, B.A. (1998) *Approximate is better than “exact” for interval estimation of binomial proportions*, The American Statistician, 52, 119-126.
- [41] Brown, L.D., Cai, T. and DasGupta, A. (2001) *Interval estimation for a binomial proportion (with discussion)*, Statistical Science, 16, 101-133.
- [42] Aitchinson, J., Dunsmore, I.R. (1975) *Statistical Prediction Analysis*, Cambridge University Press, London.
- [43] Volodin, I. (1974) *On the discrimination of gamma and Weibull distributions. Theoretical Probability Applications*, 19, 383-390.
- [44] Hager, H.W., Bain, L.J. (1970) *Inferential procedures for the generalized gamma distribution*, Journal American Statistical Association, 65, 1601-1609.

- [45] Bain, L.J., Weeks, D.L. (1965) *Tolerance limits for the generalized gamma distribution*, Journal American Statistical Association, 60, 1142-1152.
- [46] Lawless, J.F. (1980) *Inference in the generalized gamma and log gamma distributions*, Technometrics, --bf 22, 409-419.
- [47] Pham, T., Almhana, J. (1995) *The generalized gamma distribution: its hazard rate and stress-strength model*, IEEE Transactions in Reliability, 44, 392-397.